

Patent: WO 0190185-A 11 29-NOV-2001;
CURREN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES
JOURNAL

Patent: NO 0190185-A 11 29-NOV-2001;
QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES
UNIVERSITY (US)
Location/Qualifiers

```

FEATURES
SOURCE
1. .1001
/organism="Homo sapiens"
... DNA"

```

CDS

1. : 708
/2020- "unnamed protein product"

```

/protein_id=CA235343.1
/db_xref=GI:18675412
/translation="MPEIDLNTNLTTEQPVRAANFKICTIAAPYGNEGCATPLTVR
NGNIAIPACILMVAASVAAGFPLTLDLNDPSSNGIIVRSGEMTCPPMPCSTYGR
APAGGIAPEPRGGGAPAGGAGACQVNGDEGFPVRSVPRGAPLGGAPAGGACG
APPAAGFPLTCGTPPLGAGAPPLGAGAPAGNNEGPPACIYRSPAGSGARPOESTAOQ
PLGAGAPLPSASSQVHS"
                                238
PENEAISLPASASSQVHS"
                                248

```

BASE COUNT ORIGIN	261 d	DB 6;	Length 1001;
score 1001;			

Query Match	Similarity	Pred. No.	Indels	Gaps
Best Local	100.0%	1,742,227	0	0
Matches 1001; Conservative	0	Mismatches	0	

TGGGACCTGCTGACACACAGATTGGTCGAACCTTC 60
 TGGGACCTGCTGACACACAGATTGGTCGAACCTTC 60

```

Db      1 ATGCAATTGATTCGATGACGAACTTCACTG 120
OY      61 ATTAAGGAACTATTCAAGGAGCTCCATATGCTGGGATGAGCAAGCACTATTAA 120
Db      61 ATTAAGGAACTATTCAAGGAGCTCCATATGCTGGGAGAGCAAGCACTATTAA 120
OY      61 ATTAAGGAACTATTCAAGGAGCTCCATATGCTGGGAGAGCAAGCACTATTAA 180
Db      121 TTAGCTTCAGAAATGAGATGCCATTGAATTTGCCAGTTGATGTAAGCTGCTCT 180
OY      121 TTAGCTTCAGAAATGAGATGCCATTGAATTTGCCAGTTGATGTAAGCTGCTCT 180
Db      121 TTAGCTTCAGAAATGAGATGCCATTGAATTTGCCAGTTGATGTAAGCTGCT 240
OY      181 GCTGTGCCCGAGGATTTCCATTAGAACCTTAATATGACTGTTGACGCTCTAGGAAAT 240
Db      181 GCTGTGCCCGAGGATTTCCATTAGAACCTTAATATGACTGTTGACGCTCTAGGAAAT 240
OY      241 TATGTAATTAATCTGGGAAAGGAAATATGCACTCCACAGATGCTTGTGATATGTC 300
Db      241 TATGTAATTAATCTGGGAAAGGAAATATGCACTCCACAGATGCTTGTGATATGTC 300
OY      241 TATGTAATTAATCTGGGAAAGGAAATATGCACTCCACAGATGCTTGTGATATGTC 360
Db      301 TATGGGGCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 360
OY      301 TATGGGGCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 360
Db      301 TATGGGGCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 420
OY      361 GGATATGAGACCCCAACCCGTAGGAAATGAAGCCCGCTGTGGATACAGAGCTCACT 420
Db      361 GGATATGAGACCCCAACCCGTAGGAAATGAAGCCCGCTGTGGATACAGAGCTCACT 420
OY      421 GTGCGATATGAGCCCCCACTTGATATGAGACCCCACTGCGATATGAGCCCCA 480
Db      421 GTGCGATATGAGCCCCCACTTGATATGAGACCCCACTGCGATATGAGCCCCA 480
OY      421 GTGCGATATGAGCCCCCACTTGATATGAGACCCCACTGCGATATGAGCCCCA 540
Db      481 CCTGTAGATATGAGCCCACTTGGATATGGAACCCCACTTGGATATGAGCC 540
OY      481 CCTGTAGATATGAGCCCACTTGGATATGGAACCCCACTTGGATATGAGCC 540
Db      481 CCTGTAGATATGAGCCCACTTGGATATGGAACCCCACTTGGATATGAGCC 600
OY      541 CCACTCTCGGATATGAGCCCACTTGGATATGGAACCCCACTTGGATATGAGCC 600
Db      541 CCACTCTCGGATATGAGCCCACTTGGATATGGAACCCCACTTGGATATGAGCC 600
OY      601 GCTCACCTGCTGATATGAGGAGGAGCTCAGGATCTACAGAGCCAGGCTCTGAA 660
Db      601 GCTCACCTGCTGATATGAGGAGGAGCTCAGGATCTACAGAGCCAGGCTCTGAA 660
OY      601 GCTCACCTGCTGATATGAGGAGGAGCTCAGGATCTACAGAGCCAGGCTCTGAA 720
Db      661 AACGAGGCTTCTTCCCTCTGCTGCTCTCTTCAAGGCACTTCTTAAGATG 720
OY      661 AACGAGGCTTCTTCCCTCTGCTGCTCTCTTCAAGGCACTTCTTAAGATG 720
Db      661 AACGAGGCTTCTTCCCTCTGCTGCTCTCTTCAAGGCACTTCTTAAGATG 720

```

[illegible]

RESULT 2	2267 bp	mRNA	linear	PK1 04 1-22
LOCUS BC022546				
DEFINITION Homo sapiens, similar to RIKEN cDNA 4930521i23 gene, clone				
ACCESSION BC022546				
VERSION BC022546.1				
KEYWORDS				
SOURCE				
ORGANISM				

REFERENCE
Strusberg, R.
AUTHORS
TITLE
Direct Submission
Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Center Drive, Room 11A03, Bethesda, MD 20892-2590.
JOURNAL

REMARK	COMMENT
<p>NIH-MGC project</p> <p>Contact: MGC help desk</p> <p>Email: cgabs-remail.nih.gov</p> <p>Wiklos Palkovits, M.D., Ph.D. (NHGRI) & Shiraki</p>	

Tissue Procurement: Michael J. Brownstein, mjbrown@stanford.edu
 cDNA Library Preparation: (RIKEN)
 Toehiyuki and Piero Carninci
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (IHL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@plex1.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

FEATURES
SOURCE

```

RRS
source
1..2267
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26816 IMAGE:4811804"
/tissue_type="Brain, hippocampus"
/clone_lib="NIH_MGC_95"
/lab_host="DH10B"
/notes="Vector: pBluescript"
32..961
/codon_start=1
/product="Similar to RIKEN cDNA 4930521I23 gene"
CDS

```

```

/protein_id="AAH22546.1"
/db_xref="GI:18490707"
/translation="MAVQSTHTENRRGALLINGESILKRSINVELSPFGRGSNVRS
GRKTKTLPTISRIVLFTISCSISDPMLEFMPPELDLMTLVLPVPAANFKITGLQAA
PYGSGEQATKFLVFRNGDAIEPQDLMTKASAPLRRITLTDNPFSSKQIVYITBS
GNMTCPPQMRCSVIVYGAPPAGYGAPPFGYGAAPAGYGAPQNGNSGAPFVGYSASVRRG
APLPYGAAPPAGYGAPPLGYGAPPLGYGAPPLGYGAPPLGYGAPPAGNCGPPAGYTRAS
PAGGAPRHESTAAQAPENESLSPASSQVHS"

```

	Query Match	Similarity	78.2%	Score	783.2	DB	93	Length	2267	
	Best Local	Similarity	99.0%	Pred. No.	4,762	200				
	Matches	788	Conservative	0	Mismatches	8	Indels	0	Gaps	0
QY	1	ATGCCATTGATCTGATGACGAACCTCACTGTTGAACAACAGATATTTGCTGCAACTTC	60							
Db	254	ATGCCATTGATCTGATGACGAACCTCACTGTTGAACAACAGATATTTGCTGCAAACCTTC	313							
QY	61	ATTAAAGGAACTAATCAGGACGCTCCATATGGTGTCTGGAAAGGCAAGCTCACTTTTAA	120							
Db	314	ATTAAAGGAACTAATCAGGACGCTCCATATGGTGTCTGGAAAGGCAAGCTCACTTTTAA	373							
QY	121	TTAGCTTCAGAAATGAGATGCAATGCAATTGAATTGCGAGTTGATGCTGAAAGCTGCTCT	180							
Db	374	TTAGCTTCAGAAATGAGATGCAATGCAATTGAATTGCGAGTTGATGCTGAAAGCTGCTCT	433							
QY	181	GCTGTGTCGCCAGGATTTTCCACTTATGAACCTTAAATGACTGTGACTCTATAGGAATT	240							
Db	434	GCTGTGTCGCCAGGATTTTCCACTTATGAACCTTAAATGACTGTGACTCTATAGGAATT	493							
QY	241	TATGTAATTAATCTGGGAAAGGAAATATGTGACTCACAAGATGCTGTGTCAGTTATATGTC	300							
Db	494	TATGTAATTAATCTGGGAAAGGAAATATGTGACTCACAAGATGCTGTGTCAGTTATATGTC	553							
QY	301	TATGGGCCCCCACTGACAGATATGAGCCCCCACTCCCGGATACGAGCCCCCACTGCA	360							
Db	554	TATGAGCCCCCACTGACAGATATGAGCCCCCACTCCCGGATACGAGCCCCCACTGCA	613							
QY	361	GGATATGAGACCCCAACCCCTGTAGAAATGAAGCCCGCTGTGGATATACAGACCTCACT	420							
Db	614	GGATATGAGACCCCAACCCCTGTAGAAATGAAGCCCGCTGTGGATATACAGACCTCACT	673							
QY	421	GTGCGATATGAGACCCCACTCTTGATATAGGAGCCCCCACTGACAGATATGAGCCCCA	480							
Db	674	GTGCGATATGAGACCCCACTCTTGATATAGGAGCCCCCACTGACAGATATGAGCCCCA	733							
QY	481	CCTTAGGATATGAGCCCCCACTCTTGATATGAGACCCCACTCTTGATATGAGACC	540							
Db	734	CCTTAGGATATGAGCCCCCACTCTTGATATGAGACCCCACTCTTGATATGAGACC	793							
QY	541	CCACCTCCGGATATGAGCCCCCACTGACAGAAATGAAGCCCGGCTGCGGATATACGA	600							
Db	794	CCACCTCCGGATATGAGCCCCCACTGACAGAAATGAAGCCCGGCTGCGGATATACGA	853							
QY	601	GCTTCACTGCTGATATCAGAGACCGGCTCAGAAATCTACAGACAGCCCAAGCTCTGAA	660							
Db	854	GCTTCACTGCTGATATCAGAGACCGGCTCAGAAATCTACAGAGACCCCAAGCTCTGAA	913							
QY	661	AACGAGGCTTCTCTTCCCTCTGCTCTCTTCTCAGGTCCATTCTTAACTTTTAAATG	720							
Db	914	AACGAGGCTTCTCTTCCCTCTGCTCTCTTCTCAGGTCCATTCTTAACTTTTAAATG	973							
QY	721	TAAACCTTGAAGACTCACCAAGCAAGAGGTACCTTAATTTGAAGTCAAGATTAAGAGG	780							
Db	974	TAAACCTTGAAGACTCACCAAGCAAGAGGTACCTTAATTTGAAGTCAAGATTAAGAGG	1033							
QY	781	ACGACTCAGCTTAAAG	796							
Db	1034	ACGACTCAGCTTAAAG	1049							

BC022549	BC022549	2266 bp	mRNA	linear	PRI 04-FEB-2002
LOCUS					
DEFINITION	Homo sapiens, similar to RIKEN CDNA 4930521I23 gene, clone				
	MG:26828 IMAGE:4815849, mRNA, complete cds.				
ACCESSION	BC022549				
VERSION	BC022549.1	GI:18490710			
KEYWORDS	MG.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1 (bases 1 to 2266)				
AUTHORS	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-FEB-2002) National Institutes of Health, Mammalian				
	Gene Collection (MGC), Cancer Genomics Office, National Cancer				
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,				
	USA				
REMARK	NIH-MGC Project URL: http://mgc.ncl.nih.gov				
COMMENT	Contact: MGC help desk				

Tissue Procurement: Mikiros Palkovits, M.D., Ph.D.
 DNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxill.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

FEATURES	SOURCE	Location/Qualifiers
		1..2266
		/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="taxon:9606"
		/clone="MGC:26828 IMAGE:4815849"
		/tissue_type="Brain, hippocampus"
		/clone_id="NIH_MGC_95"
		/lab_host="DH10B"
		/note="Vector: pBluescript"
CDS		32..634
		/codon_start=1
		/product="Similar to RIKEN CDNA 4930521123 gene"
		/protein_id="AA422549.1"
		/db_xref="GI:18490711"
		/translation="MAVNSHTNRRGALIPNGESILKRSFNYELSPQRESGNVFS GRKQPLFTSYRVIFITSCISIDPMLSPMPPDLMTNLTVEQVPAANITKGIIOAA PYGSGEGATPDKVFRNGCAIIFPAOLMTKASAAAGRPRLRTLDNWSNGIYVITCK GICAHRCVLQGLSMERHLDMEPRHPLDTFPHLDMEPRNP"
BASE COUNT		709 a 477 c 468 g 612 t
ORIGIN		

Query Match	77.0%	Score 771.2;	DB 9;	Length 2266;
Best Local Similarity	98.9%	Prod. No. 8.2e-197;		
Matches 787;	Conservative 0;	Mismatches 8;	Indels 1;	Gaps 1;

QY	1	ATGCACTTTGATCTGATGACGAACCTCACTGTTGAACAACAGATATTGCTGCAACTTC	60
Db	254	ATGCACTTTGATCTGATGACGAACCTCACTGTTGAACAACAGATATTGCTGCAACTTC	313
QY	61	ATTAAAGGAACATATTCAGGCAAGCTCCATATGTTGGCTGGGAAGACAAAGTAATCTTTTAAA	120
Db	314	ATTAAAGGAACATATTCAGGCAAGCTCCATATGTTGGCTGGGAAGACAAAGTAATCTTTTAAA	373
QY	121	TTTACCTTTCAGAAATGAGATGCCATTTGAATTTGGCCAGTTGATGTTGAAAGCTGCTCT	180

```

Db      |||||
374  TTAGCTTCAGAAATGAGAGTCCATGTAATTCGCCATGTAATGTAAGTAAACCTGCTCT 433
Oy      |||||
181  GCGTTGGCCCGAGATTCCTCACTTGAACCTTAATATGACTGGTCACTTAATGGAAT 240
Db      |||||
434  GCTGCGCCCGAGATTCCTCACTTGAACCTTAATATGACTGGTCACTTAATGGAAT 493
Oy      |||||
241  TATGTAATTAATCTGGGAGAGGAATATGTCATCCACAGATGCTTTTCAGTTATGTC 300
Db      |||||
494  TATGTAATTAATCTGGGAGAGGAATATGTCATCCACAGATGCTTTTCAGTTATGTC 552
Oy      |||||
301  TATGGGCCCCCAGCTGAGAGATATGAGGCCCACTCCCGATACGAGACCCCACTGCA 360
Db      |||||
553  TATGAGGCCCACTGAGAGATATGAGGCCCACTCCCGATACGAGACCCCACTGCA 612
Oy      |||||
361  GATATATGAGCCCACTGAGAGATATGAGGCCCACTCCCGATACGAGACCCCACT 420
Db      |||||
613  GATATATGAGCCCACTGAGAGATATGAGGCCCACTCCCGATACGAGACCCCACT 672
Oy      |||||
421  GTGCGATATGAGCCCACTCTTGTATACGAGGCCCACTGAGAGATATGAGGCCCA 480
Db      |||||
673  GTGCGATATGAGCCCACTCTTGTATACGAGGCCCACTGAGAGATATGAGGCCCA 732
Oy      |||||
481  CCTTGAATATGAGGCCCACTCTTGTATATGAACTTCTGATATGAGGCC 540
Db      |||||
733  CCTTGAATATGAGGCCCACTCTTGTATATGAACTTCTGATATGAGGCC 792
Oy      |||||
541  CCACTCTGAGATATGAGGCCCACTCTGAGAGAAATGAGGCCCACTGAGAGATACAGA 600
Db      |||||
793  CCACTCTGAGATATGAGGCCCACTCTGAGAGAAATGAGGCCCACTGAGAGATACAGA 852
Oy      |||||
601  GCTCACTGCTGATATGAGAGCCCTCAAGATCTTACAGACAGCCAGCTCTCTGA 660
Db      |||||
853  GCTCACTGCTGATATGAGAGCCCTCAAGATCTTACAGACAGCCAGCTCTCTGA 912
Oy      |||||
661  AACGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db      |||||
913  AACGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 972
Oy      |||||
721  TAAACCTTGAAGCTCAGCAAGCAAGAGGTAACCTTAATGAGATCAGATTAAGAGG 780
Db      |||||
973  TAAACCTTGAAGCTCAGCAAGCAAGAGGTAACCTTAATGAGATCAGATTAAGAGG 1032
Oy      |||||
781  ACGACTCAGCTTAAG 796
Db      |||||
1033  ACGACTCAGCTTAAG 1048

```

```

RESULT 4
HS250D10      220895 bp      DNA      linear      PRI 05-JUN-2003
LOCUS      Human DNA sequence from clone CTA-250D10 on chromosome 22 Contains
DEFINITION the genes for SREBP2 (sterol regulatory element binding
transcription factor 2), NAGA (alpha-N-acetylgalactosaminidase), a
gene similar to neuronal-specific septin 3, a pseudogene similar to
ANT2 (adenine nucleotide translocator 2), 2 mRNAs based on ESTs, a
genomic marker D2S1178, a CA repeat polymorphism, ESTs and a Cpg
island, complete sequence.
259716      4      GI:4456457
VERSION      HTG; ANT2; Cpg Island; D2S1178; NAGA; septin 3; SREBP2.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 220895)
REFERENCE      Clark, G.
AUTHORS      Direct Submission
TITLE      Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL      Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Mar 21, 1999 this sequence version replaced gi:416439.
During sequence assembly data is compared from overlapping clones.
COMMENT

```

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMBEP; Information on the WORMBEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormbep CTA-250D10 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218.

VECTOR: pBAC108L

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

This sequence is the entire insert of clone CTA-250D10 The true left end of clone RP1-18601 is at 123979 in this sequence. The true left end of clone RP3-359316 is at 1339 in this sequence. The true right end of clone RP5-821D11 is at 23458 in this sequence. The true right end of clone RP3-359316 is at 118711 in this sequence.

FEATURES

```

source      1..220895
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /chromosome="22"
             /clone="CTA-250D10"
             /clone_11b="CTG78SK-A1"
             /size="29"
             /note="2.6 copies 5 mer ATTTG 26% conserved"
             /note="3.325"
             /note="AluX repeat: matches 1. .301 of consensus"
             /note="2.6 copies 5 mer TTGA 26% conserved"
             /note="3.338"
             /note="AluX repeat: matches 1. .301 of consensus"
             /note="2.1 copies 8 mer AGAAGACC 34% conserved"
             /note="740. .751"
             /note="2.4 copies 5 mer TTTTA 24% conserved"
             /note="1443. .1457"
             /note="7.5 copies 2 mer AC 21% conserved"
             /note="1715. .1997"
             /note="AluYo repeat: matches 5. .293 of consensus"
             /note="2103. .2378"
             /note="AluB repeat: matches 2. .296 of consensus"
             /note="2844. .2855"
             /note="12.0 copies 1 mer A 24% conserved"
             /note="complement(2884. .3033)"
             /note="MIR repeat: matches 70. .235 of consensus"
             /note="complement(3040. .3344)"
             /note="AluY repeat: matches 1. .310 of consensus"
             /note="3446. .3681"
             /note="AluM4 repeat: matches 5256. .5471 of consensus"
             /note="3746. .3837"
             /note="AluM4 repeat: matches 5615. .5707 of consensus"
             /note="3838. .4142"
             /note="AluYo repeat: matches 2. .305 of consensus"
             /note="4143. .4441"

```



```

/nc="L1M64a repeat: matches 5707. .5999 of consensus"
4442. .4747
/nc="Alusg repeat: matches 1. .292 of consensus"
repeat_region
4748. .4858
/nc="L1M64a repeat: matches 5999. .6110 of consensus"
repeat_region
4881. .5172
/nc="Alusx repeat: matches 1. .296 of consensus"
repeat_region
5418. .5429
/nc="2.4 copies 5 mer GCTGA 24% conserved"
complement(5769. .6021)
misc_feature
/nc="match: GSS: Em:A0069839"
repeat_region
6074. .6089
/nc="2.0 copies 8 mer CCCGGCTG 32% conserved"
repeat_region
6139. .6153
/nc="2.5 copies 6 mer TTTTAA 21% conserved"
repeat_region
6141. .6155
/nc="3.0 copies 5 mer TTTAA 21% conserved"
repeat_region
6366. .6572
/nc="MIR repeat: matches 5. .214 of consensus"
repeat_region
6839. .6852
/nc="2.3 copies 6 mer GGAATT 28% conserved"
repeat_region
7088. .7114
/nc="6.8 copies 4 mer TTTA 36% conserved"
repeat_region
7091. .7109
/nc="1.9 copies 10 mer ATTATTTAC 38% conserved"
repeat_region
7122. .7418
/nc="Alusx repeat: matches 1. .300 of consensus"
repeat_region
7464. .7502
/nc="19.5 copies 2 mer AT 78% conserved"
repeat_region
7502. .7524
/nc="11.5 copies 2 mer AC 46% conserved"
repeat_region
7523. .7538
/nc="2.0 copies 8 mer CAAATAT 32% conserved"
repeat_region
7535. .7563
/nc="14.5 copies 2 mer AT 49% conserved"
repeat_region
7563. .7893
/nc="Aluy repeat: matches 1. .296 of consensus"
repeat_region
7873. .7893
/nc="5.2 copies 4 mer TGT 33% conserved"
repeat_region
7914. .8200
/nc="Alub repeat: matches 1. .290 of consensus"
repeat_region
8302. .8318
/nc="AluJo/FLAM repeat: matches 1. .117 of consensus"
repeat_region
8387. .8409
/nc="11.5 copies 2 mer AC 46% conserved"
repeat_region
8483. .8498
/nc="2.0 copies 8 mer GTGTGT 32% conserved"
repeat_region
8486. .8496
/nc="2.2 copies 5 mer GTGT 22% conserved"
repeat_region
8606. .8622
/nc="2.4 copies 7 mer AGAGAA 34% conserved"
repeat_region
8767. .8989
/nc="MER15 repeat: matches 33. .268 of consensus"
repeat_region
9004. .9014
/nc="2.2 copies 5 mer CTGG 22% conserved"
repeat_region
9037. .9140
/nc="MIR repeat: matches 17. .123 of consensus"
repeat_region
9509. .9823
/nc="Alusx repeat: matches 1. .312 of consensus"
repeat_region
9872. .9883
/nc="2.4 copies 5 mer GTCT 24% conserved"
gene
10071. .49606
/nc="SRBP2"
join(<10071. .10520,11851. .12032,14129. .14275,17038. .17259,
18577. .18691,18783. .18964,20469. .20651,21182. .21363,
23956. .24232,28082. .28251,36357. .36525,38060. .38177,
40292. .40401,41889. .42021,43570. .43738,46210. .46395,
48103. .48214,48680. .49606)
/nc="SRBP2"
/product="bk250D10.1 (sterol regulatory element binding
transcription factor 2)"
/nc="sterol regulatory element binding transcription
factor 2

```

CDS

```

match: cDNA: Em:U02031 Em:U12330 Em:U00968 Em:U09103
Em:U16995 Em:U22818 Em:U22819 Em:U12329 Em:AB017337
match: B578: Em:R36312 Em:AL040138 Em:H92808 Em:AA053886"
/evidence=not_experimental
join(<10071. .10520,11851. .12032,14129. .14275,17038. .17259,
18577. .18691,18783. .18964,20469. .20651,21182. .21363,
23956. .24232,28082. .28251,36357. .36525,38060. .38177,
40292. .40401,41889. .42021,43570. .43738,46210. .46395,
48103. .48214,48680. .48900)
/gene="SRBP2"
/nc="match: protein: Sw:Q12772 Sw:Q60429 Sw:P36956
Sw:Q60416 Tr:O15352"
/codon_start=3
/evidence=not_experimental
/product="bk250D10.1 (sterol regulatory element binding
transcription factor 2)"
/protein_id="CAB4134.1"
/db_xref="GI:5578715"
/translation="MLQPVSNQVGRPRDLPSRLGSSPPGSGSSGSSGSSSS
NRRSSGAVDPSVORSPTQVTPSPSPSASQAPTLQKVSPTVPTTPATPILQ
PRPQPPQPTVLQQTVMITPTSTPTPRLIQPLTONATSQVLPQVQSLVT
SSQVQPTVITQQQVQVQAVLQTNAGTLQTLAPVQVAVQVAVQVAVQPII
KTSILVTLTKTDSVMAVQVAPALATLTPQALQVPTLVGSSGTLTTPVM
GSRKVPDKQVPGVQLBPPKSGRRTNIIKRYRSSINDKILKGLDLSLVNDAM
HSGVLRKALIDVYKYLQVNMHLROBMNMLKLANOKKILKGLDLSLVNDAM
DRQNVLLSPASDSGQAFSPYIDSPKSPFLDDAKVDEPSPVALAMVDRS
RILCVLPLCLSPFPLSTLQMGANDQHPHSGSGSVSPSGSGGMPFMMPT
LILMLVNGVIVSVFKLVHGFVPRPSPSPSPSPSPSPSPSPSPSPSPSPSP
LQTLAVLALPTSLDLAGSLSMNVIRSLQKLVMLKLVQCRATPATAG
FDEKTSARDALAYHRLHOLHITGKLPAGSASGVNHALCAVNLAECAEKIPST
LVNHLTAMGLKTRCGILGFLASFLPSRAGSLCPHSAVDSLRMLCHPLQGRF

```

Query Match 49.3%; Score 499; DB 9; Length 220895;
 Best Local Similarity 99.0%; Pred. No. 4,4e-123;
 Matches 502; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 290 CAGTTATGCTATGAGGCCCCCAGCTGACGAGATATGAGGCCCACTCCCGGATACGAG 349
DB 170003 CAGTTATGCTATGAGGCCCCCAGCTGACGAGATATGAGGCCCACTCCCGGATACGAG 170062

QY 350 CCCACCTGACGATATGAGGCCCACTGAGAAATGAGGCCCGCTGTGGATAC 409
DB 170063 CCCACCTGACGATATGAGGCCCACTGAGAAATGAGGCCCGCTGTGGATAC 170122

QY 410 GAGCTTCACCTGTGAGATATGAGGCCCACTTGGATACGAGGCCCACTGACAGAT 469
DB 170123 GAGCTTCACCTGTGAGATATGAGGCCCACTTGGATACGAGGCCCACTGACAGAT 170182

QY 470 ATGAGGCCCACTGATATGAGGCCCACTTGGATATGAGGCCCACTTGTG 529
DB 170183 ATGAGGCCCACTGATATGAGGCCCACTTGGATATGAGGCCCACTTGTG 170242

QY 530 GATATGAGGCCCACTGATATGAGGCCCACTGAGAAATGAGGCCCGCTG 589
DB 170243 GATATGAGGCCCACTGATATGAGGCCCACTGAGAAATGAGGCCCGCTG 170302

QY 590 CGGATACAGAGCTCACTGATATGAGGCCCACTGAGAAATGAGGCCCGCTG 649
DB 170303 CGGATACAGAGCTCACTGATATGAGGCCCACTGAGAAATGAGGCCCGCTG 170362

QY 650 AGGCTCTGAAAAGAGGCTTCTTCTGCTGCTCTTCTTCTGAGTCAATTTTAA 709
DB 170363 AGGCTCTGAAAAGAGGCTTCTTCTGCTGCTCTTCTTCTGAGTCAATTTTAA 170422

QY 710 CTTCTAAGATTAACCTTAACCTCAACCAAGAAAGATACCTTAATTTGAAGTCA 769
DB 170423 CTTCTAAGATTAACCTTAACCTCAACCAAGAAAGATACCTTAATTTGAAGTCA 170482

QY 770 GGATTAAGAGAGCACTCAGCTTAGAG 796
DB 170483 GGATTAAGAGAGCACTCAGCTTAGAG 170509

```

AX359658	RESULT 5
LOCUS	AX359658 1443 bp DNA linear PAT 13-FEB-2002
DEFINITION	Sequence 4 from Patent WO0190185.
ACCESSION	AX359658
VERSION	AX359658.1 GI:18675409
KEYWORDS	.
SOURCE	unidentified
ORGANISM	unclassified.
REFERENCE	1
AUTHORS	Oko, R. and Sutovsky, P.
TITLE	Pt2 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses thereof
JOURNAL	Patent: WO 0190185-A 4 29-NOV-2001; QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES UNIVERSITY (US)
FEATURES	Location/Qualifiers
source	1..1413
CDS	/organism="unidentified" /mol_type="genomic DNA" /db_xref="taxon:32644" /note="mammalia" 36..977 /note="unnamed protein product" /codon_start=1 /protein_id="CAD2336.1" /db_xref="GI:18675410" /translation="MAVNQSHTESRRGALIPSGESVLKQCDVDLCLQKFVESYLFER GTKKDTPLTYSRVVFVPSHLYNDPMLSPFMPFPGMSCTIEQIFAPNYIKGITQA FGGWGSCGFSLSPFKCGAIAEPAQMTAKSAARGLPLSVNVPPTSLIITVVA GAIVSSQTPCGPAIVIVYGPPEPVGYTQDPSYRPPGGYAGPAGYGAPPMGAGAP VGAYPRGCVGVPRGCVGVPRGCVGVPRGCVGVPRGCVGVPRGCVGVPRGCVGVPRG NEALPRAEAHSAGNTASHSMTRMQETSLEPTISS"
BASE COUNT	377 a 363 c 369 g 304 t
ORIGIN	
Query Match	43.5%; Score 435.4; DB 6; Length 1413;
Best Local Similarity	74.7%; Pred. No. 2.9e-106;
Matches 609; Conservative	0; Mismatches 181; Indels 25; Gaps 4,
Gy	1 ATGCCATTGATCTGATGACGAACCTCACTGTTGAACAACGATATTGGTCGAATTC 60
Dd	258 ATGCCGTTGGCCTGATGATGACTGCACCATTAAGAACAACAAATTTTGCCCCAACATAC 317
Gy	61 ATTAGGGAACATTCAGGAGCTCCATATGGTGGCTGGGAAGGACAGACTACTTTAAA 120
Dd	318 ATTAAAGAACCATTCAGGACCTCCAGGTGTGGCTGGGAAGGACAGACTGTTTTAAG 377
Gy	121 TTAGTCTTCAGAAATGAGATGCCATTGAATTTGGCCAGTTGATGGTGAAGCTCCTCT 180
Dd	378 TTATCTTTCAGAAAGAGAGTCCATTCGATTTGGCCACATGATGGTAAAGCTGCCCT 437
Gy	181 GCTGTTGCCGAGATTTCCACTTAGAACCTTAAATGACTGGTTCAGCTCTATGGAAAT 240
Dd	438 GCTGCTGCCAGAGGAATTCACCTTGGAAGTGAATAATTAATCTGGTTGCACATTCAGACTG 497
Gy	241 TATGTAATTACTGGGAGAGGAAT---ATGTCACTCCACAGATGCTTTGCAG----- 292
Dd	498 TACATTAATTACTGTCCAGGAGGCTGCAAGTGTCTCTCTCACACACACTTGTTCACACATAT 557
Gy	293 ----TTATGTTCTATGGGAGCCCACTTGACAGATATGAGGCCCACTTCCGAGATACGGA 348
Dd	558 CCAATTTGATCTTATGAGACCCCAACAACAGGATATACAGTCCAAACAGGGGAAATATGGA 617
Gy	349 GCCCACTGACAGATATGAGGCCCAACCCTGAAGAAATGAAGGCCGCTGTGGATAC 408
Dd	618 ACTCAACCGAAGGATATGAGACCCCAACAGGGGGAATATGAGGCCCACTTATGGGATAT 677
Gy	409 AGAGCTCACTGTGCGATATGAGGCCCAACCTCTTGAATACGAGAGCCCACCTGACAGA 468
Dd	678 GAGAGCCGCTGTGGATATGAGATCCCACTTGGGGGAATATGAGATCCCACTTGGGGGA 737

QY	469	TATGAGCCCCACCTCTGAGATATATGAGCCCACTCTTGATATGAAACCCACCTCTC	528
Db	738	TATGAGATCCCACTGGAGGATATATGAGCCCACTCTGGAGATATGAGATGCCACTGAGG	797
QY	529	GGATATGAGACCCCACTCTCGGATATATGAGCCCACTCTGAGAGAAATGAAAGCCCGCTC	588
Db	798	GGATATGATGCCCCCACTGGAGGATATATGAGCCCACTCTGAGAGATATGAGAGCCCACTCA	857
QY	589	GCGGAGATACAGAGCCTCACTGCTGATATGAGAGCCAGGCTCAGGAATCTACAGACGCC	648
Db	858	GCTGGAATATGAAAGCCCTACCCCTGCAATATGAAAGCTCATCTGCTGGAATATACAGCTGCC	917
QY	649	-----CAGGCTCTTGAAAAAGAGGCTCTCTTCCCTCTGCTCTCTCTCTCAG	696
Db	918	TCTCAGATCTATGACAGCTCAGCGAGAACTTCTTCCCACTACCTCAATCTTCTTAG	977
QY	697	GTCATTTCTTAACCTTCTTAAGATGTAAACCTTGAGACTCAGCAAGCAAGAAAGGTACCTT	756
Db	978	GTCATTTACCACTTCTCAGAGTTAACTTGTAAGACTCACCAGCAAAAG-GGCACTCT	1036
QY	757	AAAATTGAAGTCAGGATATAGAGAGACACTCAGCT	791
Db	1037	AAAATGAGTCACAGTATAGAGAGAAACCCAGGT	1071

```

RESULT 6
BX296515
LOCUS
DEFINITION
BX296515 179222 bp DNA linear HTG 26-MAY-2003
Sus scrofa clone p1gE-121D21, *** SEQUENCING IN PROGRESS ****, 6
unordered pieces.
ACCESSION
BX296515
VERSION
BX296515.6 GI:31076160
HTG; HTGS PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULFOP.
SOURCE
Sus scrofa (pig)
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 179222)
Tracy, A.
Direct Submission
Submitted (25-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
On May 26, 2003 this sequence version replaced gi:31043704.
COMMENT
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
Project Information
Center project name: B121D21
Summary Statistics
Assembly program: XGAP; version 4.5
Chemistry: Dye-terminator; 10% of reads
Consensus quality: 177502 bases at least Q40
Consensus quality: 177971 bases at least Q30
Consensus quality: 178283 bases at least Q20
Insert size: 178722; sum-of-contigs
Insert size: 167315; 12.1% error; agarose-fp
Quality coverage: 7.82x in Q20 bases; sum-of-contigs Quality
coverage: 8.59x in Q20 bases; agarose-fp
NOTE: This is a 'working draft' sequence. It currently
consists of 6 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 5850: contig of 5850 bp in length
5851 5950: gap of 100 bp
5951 74264: contig of 68314 bp in length

```

	*	74265	74364:	gap of 100 bp	
	*	74365	84655:	contig of 10299 bp in length	
	*	84657	84756:	gap of 100 bp	
	*	84757	108493:	contig of 23737 bp in length	
	*	108494	108593:	gap of 100 bp	
	*	108594	114454:	contig of 5861 bp in length	
	*	114455	114554:	gap of 100 bp	
	* 114555	179222:	contig of 64666 bp in length.		
FEATURES					
source					
		1..179222	/organism="Sus scrofa"		
			/mol_type="genomic DNA"		
			/db_xref="taxon:9823"		
			/clone_id="PigB-121D21"		
		1..5850	/clone_id="PigB"		
misc_feature			/note="assembly fragment:02838		
			fragment chain:1"		
		5951..74264			
misc_feature			/note="assembly fragment:01479		
			fragment chain:1"		
		74365..84656			
misc_feature			/note="assembly fragment:01113		
			fragment chain:1"		
		84757..108493			
misc_feature			/note="assembly fragment:00846		
			fragment chain:1"		
		108594..114454			
misc_feature			/note="assembly fragment:01058		
			fragment chain:2"		
		114555..179222			
			/note="assembly fragment:02416		
			fragment chain:2"		
BASE COUNT	44892 a	44880 c	44117 g	44831 t	502 others
ORIGIN					
Query Match	27.3%;	Score 273.6;	DB 2;	Length 179222;	
Best Local Similarity	75.0%;	Pred. No. 2,7e-62;			
Matches 372;	Conservative	0; Mismatches 114;	Indels 10;	Gaps 2;	
OY	301	TATGGGCCCCACCTCGAGGATTATGAGCCCACTTCGGGATTACGAGGCCCACTTGCA	360		
Db	55424	TACGAGCCCACCTGAAGAAGCTGAAAGCCCAACAGCAGATATGAGGCCCATTTGG	55488		
OY	361	GGATTATGAGCCCAACCCTGTAGAAATGAAGCCCGCCTGTGGATTACAGAGCCTCACCT	420		
Db	55484	GGATTATGAAGCTCACCTGGAGATTATGAGCCCTGCCTCAGAGATTGAGAGCTCACCT	55543		
OY	421	GTGCGATTATGAGCCCACTCTTGTATACGAGGCCCACTTGCAAGATTATGAGCCCA	480		
Db	55544	GAAAGACAAGAGCCCAACAGCAGATATGAGAGTCCCACTTCGGGATTATGAAGCCCA	55603		
OY	481	CCTTAGAATTATGAGCCCAACTCTTGTATATGAAACCACTCTCGGATTATGAGCC	540		
Db	55604	CCTCAAGATTATGAGCCCGCCTCAGAGATTATGAGCCTCACTGAAGATTATGAGCC	55663		
OY	541	CCAACCTCGGATTATGAGCCCACTGTGAGAAATGAAGCCCGCCTGCGGATTACAG	600		
Db	55664	CCCTCTGAGGATTATGAGCACACAAAGCTGGAATTAAGCCCACTTAAGATTATGA	55723		
OY	601	GCTCACTCTGTGATTACAGAGCCAGGCTTCAGGAATCTACAGAGCCCAAGGCTCTCTGA	660		
Db	55724	GCCCCATCTGTGAAATAAGAGCTGCTCTCACAAATCTGTGGCAAGCCAGCC-----	55787		
OY	661	AACGAGGCTCTCTTCCTCTGCGCTCCTCTTCAAGTTCACATCTTAACCTTAAGATG	720		
Db	55777	--GAGAGCTTCTCTTCCTCTTCACTCATCTTCAAGGCCATTACCACTTCTTAAGAG	55834		
OY	721	TAAACCTTGAAGCTCAACAAAGAGGTACCTTAAATTGAATGAGATTATGAGAG	780		
Db	55835	TAAACCTTGAAGATTACCAAGCAAAG-GGTACCTTAAACTGAATGATTAAGAGG	55893		
OY	781	ACGACTACGCTTAAGG	796		

```

Db      |          |||||         |           |       |
55894   AGGACTCAGCATGTG 55909

RESULT 7
BX470149
LOCUS    BX470149
DEFINITION Danio rerio clone DKEX-242K7, *** SEQUENCING IN PROGRESS ***, 50
ACCESSION BX470149
VERSION   GI:30424228
KEYWORDS HTG; HTGS PHASBI.
SOURCE    Danio rerio (zebrafish)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprindae; Danio.
1 (bases 1 to 222469)
Burton,J.
Direct Submission
Submitted (06-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 7, 2003 this sequence version replaced gi:30387077.

----- Genome Center -----
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
Project Information
Center project name: ZK24ZK7
----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 197958 bases at least Q40
Consensus quality: 206715 bases at least Q30
Consensus quality: 212351 bases at least Q20
Insert size: 217569; sum-of-contigs
Insert size: 165837; 6.2% error; agarose-fp
Quality coverage: 2.48x in Q20 bases; sum-of-contigs Quality
coverage: 3.95x in Q20 bases; agarose-fp

***** NOTE: This is a 'working draft' sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated wth the finished sequence
* as soon as it is available and the accession number will
* be prealtered.
*****
1
6656: contig of 6656 bp in length
6657        6756: gap of 100 bp
6757        10255: contig of 3499 bp in length
10256        10355: gap of 100 bp
10356        13223: contig of 2868 bp in length
13224        13323: gap of 100 bp
13324        19935: contig of 6612 bp in length
19936        20035: gap of 100 bp
20036        22597: contig of 2562 bp in length
22598        22697: gap of 100 bp
22698        28569: contig of 5872 bp in length
28570        28669: gap of 100 bp
28670        37965: contig of 9296 bp in length
37966        38065: gap of 100 bp
38066        41140: contig of 3075 bp in length
41141        41240: gap of 100 bp
41241        45108: contig of 3868 bp in length
45109        45208: gap of 100 bp
45209        47730: contig of 2522 bp in length
47731        47830: gap of 100 bp
47831        51884: contig of 4054 bp in length
51885        51984: gap of 100 bp
51985        55857: contig of 3873 bp in length

```

```

* 55858 55957: gap of 100 bp
* 55958 60498: contig of 4541 bp in length
* 60499 60598: gap of 100 bp
* 60599 63641: contig of 3043 bp in length
* 63642 63741: gap of 100 bp
* 63742 72674: contig of 8933 bp in length
* 72675 72774: gap of 100 bp
* 72775 75502: contig of 2828 bp in length
* 75503 75702: gap of 100 bp
* 75703 78063: contig of 2360 bp in length
* 78063 78162: gap of 100 bp
* 78163 84485: contig of 6323 bp in length
* 84486 84585: gap of 100 bp
* 84586 87553: contig of 2968 bp in length
* 87554 87653: gap of 100 bp
* 87654 92487: contig of 4834 bp in length
* 92488 92587: gap of 100 bp
* 92588 96104: contig of 3517 bp in length
* 96105 96204: gap of 100 bp
* 96205 101479: contig of 5275 bp in length
* 101480 101579: gap of 100 bp
* 101580 104979: contig of 3400 bp in length
* 104980 105079: gap of 100 bp
* 105080 108003: contig of 2924 bp in length
* 108004 108103: gap of 100 bp
* 108104 113151: contig of 5048 bp in length
* 113152 113251: gap of 100 bp
* 113252 116585: contig of 3334 bp in length
* 116586 116685: gap of 100 bp
* 116686 119346: contig of 2661 bp in length
* 119347 119446: gap of 100 bp
* 119447 123125: contig of 3679 bp in length
* 123126 123225: gap of 100 bp
* 123226 131366: contig of 8041 bp in length
* 131367 136193: contig of 4827 bp in length
* 136194 136293: gap of 100 bp
* 136294 139482: contig of 3189 bp in length
* 139483 139582: gap of 100 bp
* 139583 141875: contig of 2293 bp in length
* 141876 141975: gap of 100 bp
* 141976 144122: contig of 2147 bp in length
* 144123 144222: gap of 100 bp
* 144223 146968: contig of 2746 bp in length
* 146969 147068: gap of 100 bp
* 147069 149399: contig of 2331 bp in length
* 149400 149499: gap of 100 bp
* 149500 153963: contig of 4464 bp in length
* 153964 154063: gap of 100 bp
* 154064 156579: contig of 2516 bp in length
* 156580 156679: gap of 100 bp
* 156680 158928: contig of 2249 bp in length
* 158929 159028: gap of 100 bp
* 159029 164542: contig of 5514 bp in length
* 164543 164642: gap of 100 bp
* 164643 172054: contig of 7412 bp in length
* 172055 172154: gap of 100 bp
* 172155 175210: contig of 3056 bp in length
* 175211 175310: gap of 100 bp
* 175311 179379: contig of 4069 bp in length
* 179380 179479: gap of 100 bp
* 179480 182698: contig of 3219 bp in length
* 182699 182798: gap of 100 bp
* 182799 186779: contig of 3981 bp in length
* 186780 186879: gap of 100 bp
* 186880 195526: contig of 8747 bp in length
* 195527 195626: gap of 100 bp
* 195627 199396: contig of 3670 bp in length
* 199397 199496: gap of 100 bp
* 199497 206412: contig of 6916 bp in length
* 206413 209512: gap of 100 bp
* 209513 209619: contig of 3107 bp in length
* 209620 209719: gap of 100 bp

```

```

FEATURES
    source                * 209720 219083: contig of 9364 bp in length
                        * 219084 219183: gap of 100 bp
                        * 219184 222469: contig of 3286 bp in length.
                        Location/Qualifiers
                            1..222469
                                /organism="Danio rerio"
                                /mol_type="genomic DNA"
                                /db_xref="taxon:7955"
                                /clone="DKRY-242K7"
                                /clone_1b="DanioKey"
                                1..6656
                                    /note="assembly_fragment:01359
                                        fragment_chain:1"
                                6757..10255
                                    /note="assembly_fragment:01385
                                        fragment_chain:1"
                                10356..13223
                                    /note="assembly_fragment:00592
                                        fragment_chain:1"
                                13324..19935
                                    /note="assembly_fragment:00237
                                        fragment_chain:1"
                                20036..22597
                                    /note="assembly_fragment:00949
                                        fragment_chain:1"
                                22698..28569
                                    /note="assembly_fragment:00615
                                        fragment_chain:1"
                                28670..37965
                                    /note="assembly_fragment:01306
                                        fragment_chain:2"
                                38066..41140
                                    /note="assembly_fragment:00487
                                        fragment_chain:2"
                                41241..45108
                                    /note="assembly_fragment:00232
                                        fragment_chain:2"
                                45209..47730
                                    /note="assembly_fragment:00935
                                        fragment_chain:2"
                                47831..51884
                                    /note="assembly_fragment:01415
                                        fragment_chain:2"
                                51985..55857
                                    /note="assembly_fragment:00878
                                        fragment_chain:3"
                                55958..60498
                                    /note="assembly_fragment:01208
                                        fragment_chain:3"
                                60599..63641
                                    /note="assembly_fragment:01266
                                        fragment_chain:3"
                                63742..72674
                                    /note="assembly_fragment:01329
                                        fragment_chain:3"
                                72775..75502
                                    /note="assembly_fragment:01183
                                        fragment_chain:4"
                                75703..78062
                                    /note="assembly_fragment:01179
                                        fragment_chain:4"
                                78063..84485
                                    /note="assembly_fragment:01359
                                        fragment_chain:1"
                                84486..87553
                                    /note="assembly_fragment:01385
                                        fragment_chain:1"
                                87554..92487
                                    /note="assembly_fragment:00592
                                        fragment_chain:1"
                                92488..96104
                                    /note="assembly_fragment:00237
                                        fragment_chain:1"
                                96105..101479
                                    /note="assembly_fragment:00949
                                        fragment_chain:1"
                                101480..104979
                                    /note="assembly_fragment:01306
                                        fragment_chain:2"
                                104980..108003
                                    /note="assembly_fragment:00487
                                        fragment_chain:2"
                                108004..108103
                                    /note="assembly_fragment:00232
                                        fragment_chain:2"
                                108104..113151
                                    /note="assembly_fragment:00615
                                        fragment_chain:1"
                                113152..113251
                                    /note="assembly_fragment:01306
                                        fragment_chain:2"
                                113252..116585
                                    /note="assembly_fragment:00487
                                        fragment_chain:2"
                                116586..116685
                                    /note="assembly_fragment:00237
                                        fragment_chain:2"
                                116686..119346
                                    /note="assembly_fragment:00935
                                        fragment_chain:2"
                                119347..119446
                                    /note="assembly_fragment:01415
                                        fragment_chain:2"
                                119447..123125
                                    /note="assembly_fragment:00878
                                        fragment_chain:3"
                                123126..123225
                                    /note="assembly_fragment:01208
                                        fragment_chain:3"
                                123226..131366
                                    /note="assembly_fragment:01266
                                        fragment_chain:3"
                                131367..136193
                                    /note="assembly_fragment:01329
                                        fragment_chain:3"
                                136194..136293
                                    /note="assembly_fragment:01183
                                        fragment_chain:4"
                                136294..139482
                                    /note="assembly_fragment:01179
                                        fragment_chain:4"
                                139483..139582
                                    /note="assembly_fragment:01359
                                        fragment_chain:1"
                                139583..141875
                                    /note="assembly_fragment:01385
                                        fragment_chain:1"
                                141876..141975
                                    /note="assembly_fragment:00592
                                        fragment_chain:1"
                                141976..144122
                                    /note="assembly_fragment:00237
                                        fragment_chain:1"
                                144123..144222
                                    /note="assembly_fragment:00949
                                        fragment_chain:1"
                                144223..146968
                                    /note="assembly_fragment:01306
                                        fragment_chain:2"
                                146969..147068
                                    /note="assembly_fragment:00487
                                        fragment_chain:2"
                                147069..149399
                                    /note="assembly_fragment:00232
                                        fragment_chain:2"
                                149400..149499
                                    /note="assembly_fragment:00615
                                        fragment_chain:1"
                                149500..153963
                                    /note="assembly_fragment:01306
                                        fragment_chain:2"
                                153964..154063
                                    /note="assembly_fragment:00487
                                        fragment_chain:2"
                                154064..156579
                                    /note="assembly_fragment:00935
                                        fragment_chain:2"
                                156580..156679
                                    /note="assembly_fragment:01415
                                        fragment_chain:2"
                                156680..158928
                                    /note="assembly_fragment:00878
                                        fragment_chain:3"
                                158929..159028
                                    /note="assembly_fragment:01208
                                        fragment_chain:3"
                                159029..164542
                                    /note="assembly_fragment:01266
                                        fragment_chain:3"
                                164543..164642
                                    /note="assembly_fragment:01329
                                        fragment_chain:3"
                                164643..172054
                                    /note="assembly_fragment:01183
                                        fragment_chain:4"
                                172055..172154
                                    /note="assembly_fragment:01179
                                        fragment_chain:4"
                                172155..175210
                                    /note="assembly_fragment:01359
                                        fragment_chain:1"
                                175211..175310
                                    /note="assembly_fragment:01385
                                        fragment_chain:1"
                                175311..179379
                                    /note="assembly_fragment:00592
                                        fragment_chain:1"
                                179380..179479
                                    /note="assembly_fragment:00237
                                        fragment_chain:1"
                                179480..182698
                                    /note="assembly_fragment:00949
                                        fragment_chain:1"
                                182699..182798
                                    /note="assembly_fragment:01306
                                        fragment_chain:2"
                                182799..186779
                                    /note="assembly_fragment:00487
                                        fragment_chain:2"
                                186780..186879
                                    /note="assembly_fragment:00232
                                        fragment_chain:2"
                                186880..195526
                                    /note="assembly_fragment:00615
                                        fragment_chain:1"
                                195527..195626
                                    /note="assembly_fragment:01306
                                        fragment_chain:2"
                                195627..199396
                                    /note="assembly_fragment:00487
                                        fragment_chain:2"
                                199397..199496
                                    /note="assembly_fragment:00935
                                        fragment_chain:2"
                                199497..206412
                                    /note="assembly_fragment:01415
                                        fragment_chain:2"
                                206413..209512
                                    /note="assembly_fragment:00878
                                        fragment_chain:3"
                                209513..209619
                                    /note="assembly_fragment:01208
                                        fragment_chain:3"
                                209620..209719
                                    /note="assembly_fragment:01266
                                        fragment_chain:3"

```

Query Match 27.3% Score 273.6; DB 2; Length 222469;
 Best Local Similarity 75.0%; Pred. No. 2.8e-62;
 Matches 372; Conservative 0; Mismatches 114; Indels 10; Gaps 2;

```

QY 301 TATGGAGCCCACTGAGATATGAGCCCACTCCGATATGAGCCCACTCTCA 360
DB 197080 TACGAGCCCACTGAGAAAGCTCAAGCCCACTGAGATATGAGCCCACTCTG 197139
QY 361 GATATGAGCCCACTGAGAAAGCTCAAGCCCACTCTGAGATATGAGCCCACTCTCA 420
DB 197140 GATATGAGCCCACTGAGAAAGCTCAAGCCCACTCTGAGATATGAGCCCACTCTCA 197199

```

OY		421	TGCGCATATAGAGACCCCACTCTTGGATTAGAGGCCCAACTCAGAATAATVGAAGCCCCA	480
Dd		197200	GAAAAGCACGAAGCCCACACAGAGAAATATGAGGTCCACTTCGGGATATGAAAGCCCCA	197255
OY		481	CCTTAGANTATGAGACCACCCTCTTGATATAGAACCCCACTTCOGATATVGAAGCC	540
Dd		197260	CCTGCAAGATATGATAGCCCGCCTCCAAGATATVGAAGCCTCACTGAAAGATATVGAAGCC	197315
OY		541	CCACCTCTCGATATVGAAGCCCCCACTGCAGAAATAGAAAGCCCGCTCGCGGATACAGA	600
Dd		197320	CCCTTCGACAGATATVGAACAACAACAGCTGAAAAAGAACCCCACTTCAGACATATGAA	197375
OY		601	GCCTACCCGCTAGANTCAGAGAGCCAGAGCCTCAGAAATCTACAGACAGCCAGGCTCCAGAA	660
Dd		197380	GCCCAATCTGCTGGAATATGAGTGTGCTCTCACAAAATCTGTGACAGCCAGCC-----	197433
OY		661	AACGAGGCTTCTCTTCCTCTGCTGCTCTCTCTCTTCAAGTCCATTTTAACCTTCAAGATG	720
Dd		197433	--GGAGGCTTCTCTTCCTCTACCTATCTTCAAGCCCATTTTACACCTTCTTAAGAG	197490
OY		721	TTAACCTTGAAGACTCACAACGAAGAGGTACCTTAAATTAATGAACTCAGATPAAGAGG	780
Dd		197491	TAAACCTTGAAGATTCACCAAGCAAAG--GGTACCTTAAACTGAAGTCAAGATPAAGAGG	197545
OY		781	ACGACTCAGCTTAAG	796
Dd		197550	AGGACTCAGGTATGTTG	197565
RESULT #		AC113593	168425 bp DNA linear HTG 06-JUN-2002	
LOCUS		AC113593	Mus musculus clone RP23-363124, WORKING DRAFT SEQUENCE, 18 ordered pieces.	
DEFINITION		AC113593.3 GI:21327432		
ACCESSION		HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.		
VERSION		Mus musculus (house mouse)		
KEYWORDS		Mus musculus		
SOURCE		Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumetazoa; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
ORGANISM		Birtten,B., Linton,L., Nusbaum,C. and Lander,B.		
REFERENCE		Mus musculus, clone RP23-363124		
AUTHORS		Unpublished		
JOURNAL		2 (bases 1 to 168425)		
REFERENCE		Birtten,B., Linton,L., Nusbaum,C., Lander,B., All,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A., Camarata,J., Campilano,A., Chang,J., Chatato,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glade,S., Gord,S., Goylette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,M., Iliev,I., Johnson,R., Jones,C., Kamal,A., Karatas,A., Kelle,C., LaRoque,K., Lamazares,R., Landert,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Melgrim,J., Menais,L., Mihova,T., Mlenga,V., Murphy.T., Naylor.J., Nguyen.C., Nicol.R., Norbu.C., Norman,C.H., O'Connor,T., O'Donnell.P., O'Neill.D., Oliver.J., Peterson.K., Phunhang,P., Pierre.N., Pollara.V., Raymond.C., Retta,R., Rieback,M., Riley.R., Rise.C., Rogov.P., Roman.J., Roettli.M., Roy.A., Santos.R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer.B., Stange-Thomann,N., Stoianovic.N., Strausz.N., Subramanian.A., Talamo,J., Tesfaye,S., Theodore,J., Topham,K., Trevers,M., Trivas,N., Trisillo,J., Vasililev,H., Viet,R., Vo,A., Wilson,B., Wu,X., Wymn,B., Ye'W.J., Young.G., Zainoun,J., Zembeck.L., Zimmer.A. and Zody,M.		
TITLE		Direct Submission		
JOURNAL		Submitted (03-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE		3 (bases 1 to 168425)		
AUTHORS		Birtten,B., Linton,L., Nusbaum,C., Lander,B., All,A., Allen,N.,		

Anderson, S., Barina, N., Bastien, V., Bloom, T., Bogunavskiy, L.,
 Boubalger, B., Brown, A., Camarato, J., Campiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,
 Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Fark, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Horton, L., Huhne, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
 Lamarca, R., Landers, T., Lehotzky, J., Levine, R., Lindblad-Toh, K.,
 Liu, G., Maclean, C., MacDonald, P., Major, J., Marquis, N.,
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,
 Meness, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunhaph, P., Pierre, N.,
 Pollara, V., Raymond, C., Retta, R., Ribickak, W., Riley, R., Rise, C.,
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
 Schuback, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomson, N.,
 Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testfaye, S.,
 Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,
 Young, G., Zaimoun, J., Zembek, L., Zimmer, A. and Zody, W.
 Direct Submissions
 Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 6, 2002 this sequence version replaced gi:21313865.
 All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

 Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

 Project Information
 Center project name: 363.I.24
 Center Clone name: 363.I.24

 Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 162227 bases at least Q40
 Consensus quality: 164865 bases at least Q30
 Consensus quality: 165972 bases at least Q20
 Insert size: 162000; agarose-fp
 Insert size: 166725; sum-of-contigs
 Quality coverage: 8.6 in Q20 bases; agarose-fp
 Quality coverage: 8.3 in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently
 * consists of 18 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

 1 1412: contig of 1412 bp in length
 * 1413 1512: gap of 100 bp
 * 1513 3366: contig of 1854 bp in length
 * 3367 3466: gap of 100 bp
 * 3467 5542: contig of 2076 bp in length
 * 5543 5642: gap of 100 bp
 * 5643 7949: contig of 2307 bp in length
 * 7950 8049: gap of 100 bp
 * 8050 42610: contig of 34561 bp in length
 * 42611 42710: gap of 100 bp
 * 42711 45717: contig of 3007 bp in length
 * 45718 45817: gap of 100 bp
 * 45818 50106: contig of 4289 bp in length
 * 50107 50206: gap of 100 bp
 * 50207 52945: contig of 2739 bp in length
 * 52946 53045: gap of 100 bp

FEATURES

	a	c	g	t	others
BASE COUNT	44391	39572	38747	44015	1700
ORIGIN					

303 TGGGGCCCACTGCAGGATATGGAGCCCACTCCGGATACGAGCCCACTGCAGG 362

Db 66773 CACTTAGGT 66781

sequence.

SOURCE

JOURNAL Unpublished

JOURNAL S

JOURNAL OF

JOURNAL, S

JOURNAL, S

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GKFL

Center clone name: CH230-92M24

Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 242817 bases at least Q40

Consensus quality: 245517 bases at least Q30

Consensus quality: 247516 bases at least Q20

Estimated insert size: 257092; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)

NOTE: This sequence may represent more than one clone.

NOTE: This is a 'working draft' sequence. It currently

consists of 3 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 250781: contig of 250781 bp in length

250782 250881: gap of unknown length

250882 251966: contig of 1085 bp in length

251967 252066: gap of unknown length

252067 253149: contig of 1083 bp in length.

Location/Qualifiers

1. 253149

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-92M24"

misc_feature

2794..3331

/note="Clone boundary"

clone end:17

site:BCORI

end sequence: BH296484"

2829..3525

/note="Clone boundary"

clone end:17

site:BCORI

end sequence: BH296484"

misc_feature

complement(246782..247500)

/note="Clone boundary"

clone end:Sp6

site:BCORI

end sequence: BH296486"

misc_feature

64040 a 61494 c 61630 g 62235 t 3750 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 65.6%; Score 220.6; DB 2; length 253149;

Matches 322; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

301 TATGGGGCCCCCACTGAGATATGAGCCCACTCCCGATAGGAGCCCACTGCA 360

117979 TCTGATAGCGAGCGATGAGATATGATCCCTCTCTCTATACGTATACCTTATG 117920

361 GATATGAGAGCCCAAGCCCTGAGAAATGAGAGCCCGCTCTGTGATACAGAGCTGACCT 420

117919 GGCTATAGAGTCCCACTCTCTGGGTATGAGCCCACTCTGTGATACAGATCCCACT 117860

QY 421 GTGGCATATGAGAGCCCACTCTTGTATAGAGAGCCCACTGTGAGATATGAGCCCA 480

DB 117859 CTTGATATGAGAGCCCACTATATGAGATATGAGAGCCCACTCTTATATAGTATACAG 117800

QY 481 CTTTATGAGATATGAGAGCCCACTCTTGTATATGAGAGCCCACTCTGTGATATGAGCC 540

DB 117799 CTTATGAGATATGAGAGCCCACTCTTGTATATGAGAGCCCACTCTTATATAGTATACAG 117740

QY 541 CCACCTCTCGATATATGAGAGCCCACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600

DB 117739 CTTGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 117680

QY 601 GCTTACCTGTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660

DB 117679 GCATCTACCTGTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 117620

QY 661 AAGAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720

DB 117619 TTTGAGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 117560

QY 721 TTAACCTTGAAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780

DB 117559 TTAACCTTGAAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 117500

QY 781 AGCACTGAGCT 791

DB 117499 AGCACTGAGCT 117489

RESULT 11

AC132969 270171 bp DNA linear HTG 20-NOV-2002

AC132969 Rattus norvegicus clone CH230-321L20, WORKING DRAFT SEQUENCE.

AC132969 AC132969.3 GI:25139203

VERSION HTG; HTGS PHASE2; HTGS DRAFT; HTGS FULLTOP.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 270171)

REFERENCE

AUTHORS

Munzy D.Marie, Metzker M.Lee, Abramson S., Adams C., Alder J.,

Allen C., Allen H., Alebrooks S., Amin A., Anguiano D.,

Anyalebech V., Ayogaji A., Ayodeji M., Baca E., Baden H.,

Baldwin D., Bandaranaike D., Barber M., Barneshead M., Benahmed F.,

Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,

Bryant N., Buhay C., Burch P., Burrell R., Calderon E.,

Cardenas V., Carter K., Cavazos I., Caesar H., Center A.,

Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,

Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., pSouza L.,

Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,

Delgado O., Denison S., Deramo C., Ding Y., Dinh H., Divya K.,

Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Bayes K.,

Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,

Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,

Frazer C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,

Gebregiorgis B., Geer K., Gill R., Grady M., Guerra W., Guevara M.,

Gunaratne P., Haaland W., Hamill C., Hamilton C., Hamilton K.,

Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,

Hernandez R., Hines S., Hladun S.L., Hodgson A., Hognes M.,

Hollins B., Howells S., Hulyk S., Hume J., Ididibid D., Jackson A.,

Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,

Karpathy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,

Kowis C., Kraft C.L., Lebow J., Lewis J., Lewis L., Liu J.,

Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,

Lorenshew L., Loubeed H., Lozdo R.J., Lu X., Ma J.,

Maheshwari M., Mahindaratne M., Mahmood M., Malloy K., Mangum A.,

Mangum B., Mapua P., Martin K., Martin R., Martinez B.,

Mawhney S., McLeod M.P., McNeill T.Z., Meenen B.,

Milosavljevic A., Miner G., Minja B., Montemayor J., Moore S.,

Morgan M., Morris K., Morris S., Mundaes M., Murphy M., Natr L.,

Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,

ORGANISM Sus scrofa
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
TITLE 1 (bases 1 to 129624)
JOURNAL Burton, J.
COMMENT Submitted (02-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Apr 2, 2003 this sequence version replaced gi:29335441.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: b7231K18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 113740 bases at least Q40
Consensus quality: 118187 bases at least Q30
Consensus quality: 120712 bases at least Q20
Insert size: 126024; sum-of-contigs
Insert coverage: 154528; 4.6% error; agarose-fp
Quality coverage: 2.35x in Q20 bases; sum-of-contigs Quality
coverage: 3.19x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 4262: contig of 4262 bp in length
* 4263 4362: gap of 100 bp
* 4363 6991: contig of 2629 bp in length
* 6992 7091: gap of 100 bp
* 7092 9480: contig of 2389 bp in length
* 9481 9580: gap of 100 bp
* 9581 11688: contig of 2108 bp in length
* 11689 11788: gap of 100 bp
* 11789 14658: contig of 2870 bp in length
* 14659 14758: gap of 100 bp
* 14759 18440: contig of 3682 bp in length
* 18441 18540: gap of 100 bp
* 18541 20818: contig of 2278 bp in length
* 20819 20918: gap of 100 bp
* 20919 22969: contig of 2051 bp in length
* 22970 23069: gap of 100 bp
* 23070 27584: contig of 4515 bp in length
* 27585 27684: gap of 100 bp
* 27685 33211: contig of 5527 bp in length
* 33212 33311: gap of 100 bp
* 33312 35521: contig of 2210 bp in length
* 35522 35621: gap of 100 bp
* 35622 38528: contig of 2907 bp in length
* 38529 38628: gap of 100 bp
* 38629 42063: contig of 3435 bp in length
* 42064 42163: gap of 100 bp
* 42164 44164: contig of 2001 bp in length
* 44165 44264: gap of 100 bp
* 44265 47707: contig of 3443 bp in length
* 47708 47807: gap of 100 bp
* 47808 50737: contig of 2930 bp in length
* 50738 50837: gap of 100 bp
* 50838 57066: contig of 6229 bp in length
* 57067 57166: gap of 100 bp
* 57167 59978: contig of 2812 bp in length
* 59979 60078: gap of 100 bp
* 60079 62560: contig of 2482 bp in length

FEATURES
source * 62561 62660: gap of 100 bp
* 62661 64963: contig of 2303 bp in length
* 64964 65064: gap of 100 bp
* 65064 67192: contig of 2128 bp in length
* 67192 67292: gap of 100 bp
* 67292 72006: contig of 4714 bp in length
* 72006 72106: gap of 100 bp
* 72106 76596: contig of 4490 bp in length
* 76596 76696: gap of 100 bp
* 76696 79826: contig of 3130 bp in length
* 79826 79926: gap of 100 bp
* 79926 83540: contig of 3615 bp in length
* 83541 83640: gap of 100 bp
* 83641 86276: contig of 2636 bp in length
* 86276 86377: gap of 100 bp
* 86377 94620: contig of 8244 bp in length
* 94621 94720: gap of 100 bp
* 94721 97520: contig of 2800 bp in length
* 97521 97620: gap of 100 bp
* 97621 99784: contig of 2164 bp in length
* 99785 99885: gap of 100 bp
* 99885 102112: contig of 2228 bp in length
* 102113 102212: gap of 100 bp
* 102213 104242: contig of 2030 bp in length
* 104243 104342: gap of 100 bp
* 104343 110710: contig of 6368 bp in length
* 110711 110810: gap of 100 bp
* 110811 113167: contig of 2257 bp in length
* 113168 11367: gap of 100 bp
* 11368 11367: gap of 100 bp
* 11368 11505: contig of 2038 bp in length
* 11506 11505: gap of 100 bp
* 11506 121874: contig of 6269 bp in length
* 121875 121974: gap of 100 bp
* 121975 127096: contig of 5122 bp in length
* 127097 127196: gap of 100 bp
* 127197 129624: contig of 2428 bp in length.
Location/Qualifiers
1..129624
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9823"
/clone="P198-231K18"
/clone_1fb="P198"
1..4262
/note="assembly fragment:00361
fragment_chain:1"
4363..6991
/note="assembly fragment:00108
fragment_chain:1"
7092..9480
/note="assembly fragment:00460
fragment_chain:1"
9581..11688
/note="assembly fragment:00959
fragment_chain:1"
11789..14658
/note="assembly fragment:00995
fragment_chain:2"
14759..18440
/note="assembly fragment:00608
fragment_chain:2"
18541..20818
/note="assembly fragment:01101
fragment_chain:2"
20919..22969
/note="assembly fragment:00996
fragment_chain:2"
23070..27584
/note="assembly fragment:00217
fragment_chain:3"
27685..33211
/note="assembly fragment:00019
fragment_chain:3"


```

misc_feature      clone_end:Sp6"
                  4245_4594
                  /note="clone boundary
                  clone_end:Sp6
                  site:BCORI
misc_feature      end sequence:BH364165"
                  complement(230311..230677)
                  /note="clone boundary
                  clone_end:T7
                  site:BCORI
misc_feature      end sequence:BH364164"
                  23134..23298
                  /note="xgs_end_extension
                  clone_end:T7"

BASE COUNT      59674 a 53524 c 55897 g 56470 t 9816 others
ORIGIN
Query Match      11.8%; Score 118.2; DB 2; Length 23581;
Best Local Similarity 54.2%; Pred. No.2,3e-20;
Matches 240; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY      293 TTATGTCTATGAGGCCCCCACTGCAGATATGAGGCCCACTCCCGATACGAGCCC 352
DB      141802 TTCTATTATACATGAAACCTTCTCTATTATACATGAGACCTCTCTATTATACATGAGCAC 141743

QY      353 CACCTGAGATATGAGCCCAACCGTAGAATGAGGCCCGCTGTGGATACAGAG 412
DB      141742 CTCCTTATTAATGAGGACCTTCTATTATACATGAGACCTTCTCTATTATACATGAGAG 141683

QY      413 CTCACCTGCGATATGAGGCCCACTCTGTGATACGAGGCCCACTGCAGATATG 472
DB      141682 CACCTCTTATTATACATGAGACCTTCTATTATACATGAGACCTCTCTATTATACATG 141623

QY      473 GAGCCCCACCTCTAGATATGAGGCCCACTCTGTGATATGAGACCCCACTCTCGAT 532
DB      141622 GATCACCCTCTTATTATACATGAGACCTTCTCTATTATACATGAGACCTTCTCTATTAC 141563

QY      533 ATGAGAGCCCACTCTCGATATGAGGCCCACTCTGCAGAAATGAGGCCCGCTCGAG 592
DB      141562 ATGATACACCTCTTATTATACATGAGGACCTTCTCTATTATACATGATCACTCTCTTAT 141503

QY      593 GATACAGAGCCTCACCGCTGATATGAGGACGAGGCGCTCGAATACAGAGCCAG 652
DB      141502 TATACATGAGACCTCTCTATTATACATGATACCTCTTATTATACATGAGACCTTCTC 141443

QY      653 CTCCTGAAAACGAGGCTTCTCTCCCTCTGCTCTCTCTCTGAGGTCATTCTTAACCTT 712
DB      141442 TATTACACAGAGCACTTCTCTATTATACAGGAGACCTCTCTATTATACATGAGACCTT 141383

QY      713 CTAAAGATTAACCTTGAGACT 735
DB      141382 CTCATTATTAATACATGAGCACT 141360

RESULT 15
AC108000      167587 bp      DNA      linear      PRI 21-MAY-2002
LOCUS      Homo sapiens chromosome 15, clone CTD-2116G1, complete sequence.
DEFINITION      AC108000
ACCESSION      AC108000
VERSION      AC108000.5 GI:21039854
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 167587)
REFERENCE      Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
JOURNAL      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Homo sapiens chromosome 15, clone CTD-2116G1
REFERENCE      Unpublished
2 (bases 1 to 167587)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Brown,A., Camarata,J., Campopiano,A., Chang,D., Chazaro,B.,

```

```

TITLE
JOURNAL
AUTHORS
REFERENCE
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 167587)
Submitted (28-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 167587)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barta,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Gardyna,S., Gird,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatae,A., Kelle,C., Lacroque,K., Lamazares,R.,
Lander,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marcuis,N., Matthews,C., McCarthy,M.,
McGowan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,D., Nguyen,C., Nicol,R., Nordu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnpback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 21, 2002 this sequence version replaced gi:20336146.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L24580

Center clone name: 2116_G_1

FEATURES

```

SOURCE
1..167587
    location/Qualifiers
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="15"
    /map="15"
    /clone="CTD-2116G1"
    /clone_id="CTD1 Human BAC"
repeat_region
    complement(11..144)
    /rpt_family="L2"
repeat_region
    complement(145..356)
    /rpt_family="Alusg/x"
    complement(357..1109)
    /rpt_family="L2"
repeat_region
    3713..3828
    /rpt_family="L1ME"
    complement(3834..4006)
    /rpt_family="MIR"
repeat_region
    4574..4603
    /rpt_family="(GAAA)n"
    complement(4976..5167)
    /rpt_family="MIR"
repeat_region
    5639..5775
    /rpt_family="L1"
    /rpt_family="L1"
repeat_region
    6745..6766
    /rpt_family="AT_rich"
    complement(8290..8370)
    /rpt_family="MIR3"
repeat_region
    9077..9288
    /rpt_family="MIR"
    /rpt_family="MIR"
repeat_region
    9496..10838
    /rpt_family="L1MD1"
    10839..11162
    /rpt_family="Alusx"
    11163..12229
    /rpt_family="L1MD1"
    complement(13118..13414)
    /rpt_family="AluY"
repeat_region
    13667..14061
    /rpt_family="L2"
repeat_region
    14366..14395
    /rpt_family="AT_rich"
    14748..15052
    /rpt_family="Alusx"
repeat_region
    complement(15603..15768)
    /rpt_family="MIR"
repeat_region
    complement(15878..15932)
    /rpt_family="L2"
repeat_region
    complement(15988..16069)
    /rpt_family="MER81"
repeat_region
    complement(16309..16442)
    /rpt_family="MER46A"
repeat_region
    complement(16557..16730)
    /rpt_family="MIR"
repeat_region
    complement(16746..16830)
    /rpt_family="MIR"
    16845..16939
    /rpt_family="MER117"

```

```

repeat_region complement(17267..17436)
    /rpt_family="MIR3"
repeat_region complement(17914..18014)
    /rpt_family="MER103"
repeat_region complement(19124..19415)
    /rpt_family="AluDo"
repeat_region complement(19571..19636)
    /rpt_family="L2"
    19637..19926
    /rpt_family="AluDo"
repeat_region complement(19927..19984)
    /rpt_family="L2"
repeat_region complement(20449..21788)
    /rpt_family="L2"
    21985..22144
    /rpt_family="MIR"
repeat_region 22728..22838
    /rpt_family="Char11e8"
repeat_region 23014..23044
    /rpt_family="AT_rich"
    23140..23256
    /rpt_family="FLAM_C"
    /rpt_family="FLAM_C"
repeat_region 23265..23350
    /rpt_family="(TA)n"
repeat_region 23357..23381
    /rpt_family="AT_rich"
repeat_region 24508..25206
    /rpt_family="Alusx"
    complement(25805..25979)
    /rpt_family="L1ME4A"
repeat_region complement(26043..26262)
    /rpt_family="L1ME4A"
repeat_region complement(26861..26993)
    /rpt_family="FLAM_C"
repeat_region complement(27023..27317)
    /rpt_family="Alusx"
repeat_region 27952..28300

```

Query Match

11.44; Score 113.8; DB 9; Length 167587;
 Best Local Similarity 55.18; Pred. No. 3,4e-19;
 Matches 223; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

```

QY 297 TGTCTATGAGGAGCCCACTGAGATATGAGAGCCCACTCCGAGATACGAGCCCAACC 356
    |||||
DB 119174 TGTTCACACAGCTCTCTCTCTGTTACAGTACCTCTCTCTGTTACACAGCTCTCC 119233
    |||||
QY 357 TGCAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCTTC 416
    |||||
DB 119234 TCTGTTACAGTACCTCTCTCTCTGTTACAGTACCTCTCTCTGTTACAGTACCTTC 119293
    |||||
QY 417 ACTGTGAGATATGAGAGCCCACTCTTGTATGAGAGCCCACTGAGATATGAGAGC 476
    |||||
DB 119294 TCTCTGTTACACAGCTCTCTCTCTGTTACAGTACCTCTCTCTGTTACAGAGC 119353
    |||||
QY 477 CCCACCTCTGAGATATGAGAGCCCACTCTTGTATGAGAGCCCACTCTCTGAGATATG 536
    |||||
DB 119354 CTCTCTCTCTGTTACAGTACCTCTCTCTCTGTTACAGTACCTCTCTCTGTTACAGT 119413
    |||||
QY 537 AGCCCACTCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGC 596
    |||||
DB 119414 AGCTCTCTCTCTGTTACAGTACCTCTCTCTCTGTTACAGTACCTCTCTCTGTTAC 119473
    |||||
QY 597 CAGAGCTCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGC 656
    |||||
DB 119474 CACAGCTCTCTCTCTGTTACAGTACCTCTCTCTCTGTTACAGAGCTCTCTCTCTGT 119533
    |||||
QY 657 TGAAGAAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 701
    |||||
DB 119534 TCATGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 119578
    |||||

```

Search completed: December 16, 2003, 15:10:18
 Job time : 3730.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 05:54:30 ; Search time 267.133 Seconds

10115.338 Million cell updates/sec

Title: US-09-864-291-11

Sequence: 1 atgccatttgatctgatgac.....attacctcatgtgtcaggtag 1001

Scoring table: IDENTITY_NUC

[illegible][illegible]

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_19Jun03: *

1:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1980.DAT *
2:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1981.DAT *
3:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1982.DAT *
4:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1983.DAT *
5:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1984.DAT *
6:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1985.DAT *
7:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1986.DAT *
8:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1987.DAT *
9:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1988.DAT *
10:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1989.DAT *
11:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1990.DAT *
12:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1991.DAT *
13:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1992.DAT *
14:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1993.DAT *
15:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1994.DAT *
16:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1995.DAT *
17:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1996.DAT *
18:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1997.DAT *
19:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1998.DAT *
20:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA1999.DAT *
21:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA2000.DAT *
22:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA2001A.DAT *
23:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA2001B.DAT *
24:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA2002.DAT *
25:	/SIDS1/gcgdata/geneseq/geneseqm-emb1/NA2003.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	100.1	100.0	100.1	24	AA520602	DNA encoding human
2	49.9	49.9	220895	24	AB844998	human cdna difference
3	44.6	44.6	467	22	AA101229	Human reproductive
4	44.6	44.6	467	23	AB996868	Human testicular a
5	43.5	43.5	1413	24	AA520601	DNA encoding bovine
6	43.4	43.4	436	22	AB467850	Human foetal liver
7	43.4	43.4	436	22	AAK42003	Human bone marrow
8	43.4	43.4	436	22	AA148070	Probe #16756 used

9	434.4	43.4	43.6	24	ABSL6034	Human genome-deriv
10	375.4	37.5	7099	24	AL048862	Human reproductive
11	375.4	37.5	7099	23	ABSL97776	Human testicular a
12	173.6	17.3	894	23	ABSL70582	DNA encoding novel
13	130.4	13.0	471	22	ABAS5246	Human foetal liver
14	130.4	13.0	471	22	AAK82957	Human bone marrow
15	130.4	13.0	471	22	AAI34906	Probe #3592 used t
16	130.4	13.0	471	22	ABSL03491	Human genome-deriv
17	110.2	11.0	321	22	AAAL01355	Human reproductive
18	110.2	11.0	321	23	ABSL6808	Human testicular a
19	110.2	11.0	5769	22	AL048864	Human reproductive
20	110.2	11.0	5769	23	ABSL97778	Human testicular a
21	104.2	10.4	2772	23	AAAS72787	DNA encoding novel
22	104.2	10.4	2772	23	AAAS90525	DNA encoding novel
23	102.6	10.2	2243	22	ABSL08657	Human extensin hom
24	100	10.0	1824	23	AAAS81488	DNA encoding novel
25	100	10.0	2850	23	AAAS79695	DNA encoding novel
26	93.2	9.3	291	22	ABSL97777	Human reproductive
27	93.2	9.3	291	23	AAAL04883	Human testicular a
28	89.4	8.9	1984	22	ABSL45936	Human breast cell
29	89.4	8.9	1984	22	ABSL45641	Human foetal liver
30	89.4	8.9	1984	22	ABSL60096	Probe #4562 for ge
31	89.4	8.9	1984	22	AAK046522	Human brain expres
32	89.4	8.9	1984	22	AAK30133	Human bone marrow
33	89.4	8.9	1984	22	AAI14731	Probe #4664 for ge
34	89.4	8.9	1984	22	AAI136097	Probe #4783 used t
35	89.4	8.9	1984	22	AAI04535	Probe #4526 used t
36	89.4	8.9	1984	23	ABSL29782	Human liver single
37	89.4	8.9	1984	24	ABSL07714	Human genome-deriv
38	87	8.7	543	22	ABSL05054	Human breast cell
39	87	8.7	543	22	ABSL69042	Human foetal liver
40	87	8.7	543	22	ABSL35988	Probe #14454 for g
41	87	8.7	543	22	AAK17357	Human brain express
42	87	8.7	543	22	AAK13154	Human bone marrow
43	87	8.7	543	22	AAI33925	Probe #13858 for g
44	87	8.7	543	22	AAI49232	Probe #17918 used t
45	87	8.7	543	22	AAI09525	Probe #9516 used t

ALIGNMENTS

```

RESULT 1
AAS20602
ID      AAS20602 standard; cDNA; 1001 BP

```

AC AAS20602

DT 09-APR-2002 (first entry)

DE DNA encoding human testicular WW domain binding protein (htcWBP)

KW Testicular WW domain binding protein; WBPH; perinuclear theca 32.

KW globozoospermy; spermiogenesis; spermatozoa; tyrosine kinase; c-Yes;

XX

XX

```
PT primer_bind 1..18
```

PT /note= "Blinds forward primer AAS20605"

```
/*tag= b
```

/note= "Human testicular WW domain binding protein"

```
/*tag= c
```

2000

PD 29-NOV-2001.
 XX
 PF 25-MAY-2001; 2001MO-CA00738.
 XX
 PR 25-MAY-2000; 2000CA-2307228.
 PR 25-MAY-2000; 2000US-206979P.
 XX
 PA (TOOH) UNIV QUEBENS KINGSTON.
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 XX
 PI Olo R, Sutovsky P;
 XX
 DR WPI; 2002-097644/13.
 DR P-PSDB; AAU74610.
 XX
 PT Isolated perinuclear theca 32 polypeptide that interacts with activated
 PT tyrosine kinase c-Yes, for enhancing fertility, treating/diagnosing
 PT diminished fertility and abnormal spermiogenesis and for providing
 PT contraception -
 XX
 PS Claim 62; Fig 4B; 103pp; English.
 XX
 CC The invention describes an isolated perinuclear theca 32 (PT32)
 CC polypeptide (I) which interacts with tyrosine kinase c-Yes. (I) is
 CC useful for: enhancing fertility in a mammal; treating globozoosperm; by
 CC expressing (I) in spermatozoa; inhibiting fertilization, by introducing
 CC (I) or its antigenic fragment into a mammal to elicit an immune
 CC response; enhancing the ability of round spermatids to activate oocytes;
 CC treating or diagnosing diminished fertility and abnormal spermiogenesis;
 CC in providing contraception; identifying contraceptive and
 CC fertility-enhancing agents. The polynucleotide is useful for producing
 CC (I) by recombinant techniques, as vaccine, as diagnostic reagents, and
 CC for chromosome identification. An antibody against (I) is useful in
 CC immunological assays, in immunoneutralization methods, to identify cells
 CC expressing (I), and to purify (I) by affinity chromatography. A
 CC transgenic animal is useful as an animal model for studying human
 CC fertility and reproductive biology, and for screening compounds to
 CC identify modulators of oocyte activation. The use of (I) prevents the
 CC entry of components which are detrimental to embryonic development into
 CC the oocyte during oocyte activation with crude sperm extract and avoids
 CC the propagation of viruses such as HIV (human immunodeficiency virus) and
 CC SIV (simian immunodeficiency virus) carried in the sperm. This sequence
 CC encodes the human testicular WW domain binding protein (htbWP), described
 CC in the method of the invention.
 CC
 XX
 SQ Sequence 1001 BP; 261 A; 254 C; 248 G; 238 T; 0 other;
 Query Match 100.0%; Score 1001; DB 24; Length 1001;
 Best Local Similarity 100.0%; Pred. No. 5.3e-290;
 Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGCAATTTGATCTGATGAGCAACCTGCTGTGGAACAACAGATTGCTGCAAACTTC 60
 DB 1 ATGCAATTTGATCTGATGAGCAACCTGCTGTGGAACAACAGATTGCTGCAAACTTC 60
 QY 61 ATTAAGGGAACATTTTCAGGCACTCCATATGCTGCTGCGAAGAGCAAGCTACTTTAAA 120
 DB 61 ATTAAGGGAACATTTTCAGGCACTCCATATGCTGCTGCGAAGAGCAAGCTACTTTAAA 120
 QY 121 TTGCTCTTGAGAAATGAGATGCAATTTGCTCCAGTTGATGATGGAAGCTGCTCT 180
 DB 121 TTGCTCTTGAGAAATGAGATGCAATTTGCTCCAGTTGATGATGGAAGCTGCTCT 180
 QY 181 GCTGTTGCCGAGGATTTTCACCTTAAGACCTTAATGATGCTTACAGCTCTATGGAATT 240
 DB 181 GCTGTTGCCGAGGATTTTCACCTTAAGACCTTAATGATGCTTACAGCTCTATGGAATT 240
 QY 241 TATGTAATTAATGCTGGGAAAGGGAATATGATGCACTTCACAGATGCTTGTGATTTATGTC 300
 DB 241 TATGTAATTAATGCTGGGAAAGGGAATATGATGCACTTCACAGATGCTTGTGATTTATGTC 300
 QY 301 TATGAGGAGCCCACTGAGAGATATGAGCCCACTCCGATTAAGAGCCCACTGCA 360
 DB 301 TATGAGGAGCCCACTGAGAGATATGAGCCCACTCCGATTAAGAGCCCACTGCA 360

DB 301 TATGAGGAGCCCACTGAGAGATATGAGCCCACTCCGATTAAGAGCCCACTGCA 360
 QY 361 GGATATGAGCCCACTGAGAGATATGAGCCCGCTGTGAGATACAGAGCTCACT 420
 DB 361 GGATATGAGCCCACTGAGAGATATGAGCCCGCTGTGAGATACAGAGCTCACT 420
 QY 421 GTGCGATATGAGCCCACTGAGATATGAGAGCCCACTGAGAGATATGAGCCCA 480
 DB 421 GTGCGATATGAGCCCACTGAGATATGAGAGCCCACTGAGAGATATGAGCCCA 480
 QY 481 CCTGATGATATGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCC 540
 DB 481 CCTGATGATATGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCC 540
 QY 541 CCACTCTGAGATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGCGAGATACAG 600
 DB 541 CCACTCTGAGATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGCGAGATACAG 600
 QY 601 GCTGACCTGCTGAGATACAGAGCCGCTGAGAGATATGAGAGCCCGCTGCTGAA 660
 DB 601 GCTGACCTGCTGAGATACAGAGCCGCTGAGAGATATGAGAGCCCGCTGCTGAA 660
 QY 661 AACGAGGCTTCTCTTCCCTGCTGCTCTCTCTCAAGTCCATTTAACTTAAGATG 720
 DB 661 AACGAGGCTTCTCTTCCCTGCTGCTCTCTCTCAAGTCCATTTAACTTAAGATG 720
 QY 721 TAACTTTGAGACTCAACCAAGCAAGATGACCTTAATTAAGTCAAGATTAAGAGG 780
 DB 721 TAACTTTGAGACTCAACCAAGCAAGATGACCTTAATTAAGTCAAGATTAAGAGG 780
 QY 781 AGACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
 DB 781 AGACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
 QY 841 CCATGTTCTCAAGAGCACTTCTTCAAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 900
 DB 841 CCATGTTCTCAAGAGCACTTCTTCAAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 900
 QY 901 GAAAGAGAGAGAGACTCAACCACTGAGAGAGAGATGATGATGATGATGATGATGATG 960
 DB 901 GAAAGAGAGAGAGACTCAACCACTGAGAGAGAGATGATGATGATGATGATGATGATG 960
 QY 961 AGTGTCTGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1001
 DB 961 AGTGTCTGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1001
 RESULT 2
 ABK84798
 ID ABK84798 standard; cDNA; 220895 BP.
 XX
 AC ABK84798;
 XX
 DT 14-AUG-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #1369.
 XX
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 XX
 OS Homo sapiens.
 XX
 PN W0200228999-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 03-OCT-2001; 2001MO-US30821.

XX 03-OCT-2000; 2000US-237189P.
PR (GENE-) GENE LOGIC INC.
PA
XX Beazer-Barclay Y, Weisman SM, Yamaga S, Vockley J;
PI MPI: 2002-435328/46.
XX
DR
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
PS Claim 1; SEQ ID No 1369; 114pp; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection and
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 220895 BP; 53760 A; 54808 C; 55831 G; 56496 T; 0 other;
SQ
Query Match 49.9%; Score 499; DB 24; Length 220895;
Best Local Similarity 99.0%; Pred. No. 5, 6e-138;
Matches 502; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 290 CAGTATTCCTATGAGGCGCCCACTGAGATATGAGAGCCCACTCCCGATACGAG 349
DB 170003 CAGTATTCCTATGAGGCGCCCACTGAGATATGAGAGCCCACTCCCGATACGAG 170062
QY 350 CCCCACCTGACGATATGAGAGCCCACTGAGAAATGAGAGCCGCTGAGATACA 409
DB 170063 CCCCACCTGACGATATGAGAGCCCACTGAGAAATGAGAGCCGCTGAGATACA 170122
QY 410 GAGCTCCTGACGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAT 469
DB 170123 GAGCTCCTGACGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAT 170182
QY 470 ATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAT 529

DB 170183 ATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAG 170242
QY 530 GATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAAATGAGAGCCGCTG 589
DB 170243 GATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAAATGAGAGCCGCTG 170302
QY 590 CGGATACAGAGCCCTCACTGCTGATGAGAGAGCCGCTCAGGAATCTACAGAGCC 649
DB 170303 CGGATACAGAGCCCTCACTGCTGATGAGAGAGCCGCTCAGGAATCTACAGAGCC 170362
QY 650 AGGCTCTGAAAACGAGGCTTCTTCCCTGCTCTCTTCTGAGTCAATCTTAAC 709
DB 170363 AGGCTCTGAAAACGAGGCTTCTTCCCTGCTCTCTTCTGAGTCAATCTTAAC 170422
QY 710 CTTCTAATATTAACCTTGAAGACTACCAAGCAAAAGGTACCTTAATGAATCA 769
DB 170423 CTTCTAATATTAACCTTGAAGACTACCAAGCAAAAGGTACCTTAATGAATCA 170482
QY 770 GGATTAAGAGAGCAGCTCAGCTTAGAG 796
DB 170483 GGATTAAGAGAGCAGCTCAGCTTAGAG 170509
RESULT 3
ID AAL01229 standard; cDNA, 467 BP.
XX AAL01229;
AC AAL01229;
XX
XX 21-NOV-2001 (first entry)
DT
XX
DS Human reproductive system related antigen cDNA SEQ ID NO: 1230.
XX
XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ss.
XX Homo sapiens.
XX
XX WO200155320-A2.
PD
XX
XX 02-AUG-2001.
PF
XX 17-JAN-2001; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214836.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216447.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 26-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225457.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.

```

PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 13-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241888.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.

```

```

PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
DR P-PSDB; AAM955259.
XX
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
XX 1s used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 1; SEQ ID NO 1230; 1297bp + Sequence listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a coding sequence of the
CC invention.
XX
SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;

Query Match 44.6%; Score 446.2; DB 22; Length 467;
Best Local Similarity 98.5%; Pred. No. 1.5e-123;
Matches 459; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 129 CAGAAATGAGATGACATTTGATTTGCCAGTTGATGAGAAAGCTGCTGCTGTC 188
DB 2 CAGAAATGAGATGACATTTGATTTGCCAGTTGATGAGAAAGCTGCTGCTGTC 61
QY 189 CCGAGATTTTCACCTTAGACCTTAATGACTGCTGCTGCTGCTGCTGCTGCTGCT 248
DB 62 CCGAGATTTTCACCTTAGACCTTAATGACTGCTGCTGCTGCTGCTGCTGCTGCT 121
QY 249 TACTGGGGAAGGGAATATATGCACTGCACAGATGCTTGTTCATTATTTGATGGGC 308
DB 122 TACTGGGGAAGGGAATATATGCACTGCACAGATGCTTGTTCATTATTTGATGGGC 181
QY 309 CCACCTGAGATATGAGCCCACTCCGGATACGAGAGCCCACTCCGAGATATGG 368

```

Db 182 CCCACTGACAGATATGAGCCCACTCCGGATACGAGAGCCCACTGAGATATGG 241
QY 369 AGCCCAACCCCTAGAAATGAGAGCCCGCTGTGGATACAGAGCTCACTGTGGCTA 428
Db 242 AGCCCAACCCCTAGAAATGAGAGCCCGCTGTGGATACAGAGCTCACTGTGGCTA 301
QY 429 TGAAGCCCACTCTTGATAGAGAGCCCACTGACGATATGAGAGCCCACTCTAGG 488
Db 302 TGAAGCCCACTCTTGATAGAGAGCCCACTGACGATATGAGAGCCCACTCTAGG 361
QY 489 ATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTC 547
Db 362 ATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTGATATGAGAGCCCACTCT 421
QY 548 TCGATATGAGAGCCCACTGACGAGAAATGAGAGCCCGCTGGCGG 593
Db 422 TCGATATGAGAGCCCACTGACGAGAAATGAGAGCCCGCTGGCGG 467

RESULT 4
ABL96688
ID ABL96688 standard; cDNA; 467 BP.
XX
AC ABL96688;
XX
DT 21-JUN-2002 (first entry)
XX
DE Human testicular antigen encoding cDNA SEQ ID NO: 356.
XX
KM Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KM reproductive system disorder; urinary system disorder; gene therapy;
KM cardiovascular disorder; respiratory disorder; neurological disorder;
KM gastrointestinal disease; infection; cytostatic; gene; ss.
XX
OS Homo sapiens.
PN WC00155317-A2.
PD 02-AUG-2001.
XX
PP 17-JAN-2001; 2001WO-US01329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 11-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236357.
PR 29-SEP-2000; 2000US-0236358.
PR 29-SEP-2000; 2000US-0236359.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240660.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246602.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 03-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-483232/52.
 XX
 PT Nucleic acids encoding 973 human testicular antigen polypeptides,
 XX useful for preventing, diagnosing and/or treating testicular cancer -
 XX
 PS Claim 1, SEQ ID NO 356; 766bp; English.
 XX
 CC The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer,
 CC especially testicular cancers. The present sequence is a cDNA of the
 CC invention.
 XX
 SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;

	Query Match	Similarity	44.6%	Score	446.2	DB 23	Length	467
	Best Local	Similarity	98.5%	Pred. No.	1.5e-123			
	Matches	459	Conservative	1	Mismatches	5	Indels	1
QY	129	CAGAAATGAGATGCGCATTTGAAATTTGCCAGATTGATGTGAAAGCTGCTGTGCTGTGC						188
DB	2	CAGAAATGAGAGGTGCATTTGAATTTGCCAGATTGATGTGAAAGCTGCTGTGCTGTGC						61
QY	189	CCGAGATTTTCCACTTAGAAGCTTAAATGACGTTCAGCTCATATGGAAATTATGTAT						248
DB	62	CCGAGATTTTCCACTTAGAAGCTTAAATGACGTTCAGCTCATATGGAAATTATGTAT						121
QY	249	TACTGGGAAAGGAATATGTGACTCCACAGATGCTCTGTTCAGTTATTTGTCTATGGGAC						308
DB	122	TACTGGGAAAGGAATATGTGACTCCACAGATGCTCTGTTCAGTTATTTGTCTATGGGAC						181
QY	309	CCCACTTGACAGATATGAGGCCCACTTCCGGATACGAGCCCCCACTGTCAGATATGG						368
DB	182	CCCACTTGACAGATATGAGGCCCACTTCCGGATACGAGCCCCCACTGTCAGATATGG						241

QY	369	AGCCCAACCCGTATGGAAATGAAGGCGCGCTGTGGGAATACAGAGCTCAACCTGTGCGATA	428
DB	242	AGCCCAACCCGTATGGAAATGAAGGCGCGCTGTGGGAATACAGAGCTCAACCTGTGCGATA	301
QY	429	TGAGAGCCCCACCTCTTGATATCGAGGCCCACTTGCAAGATATGAGAGCCCCACCTTAGG	488
DB	302	TGAGAGCCCCACCTCTTGATATCGAGGCCCACTTGCAAGATATGAGAGCCCCACCTTAGG	361
QY	489	ATATGAGAGCCCCACCTCTTGATATGAGAGCCCCACCTCTTGATATGAGAGCCCCACCTC	547
DB	362	ATATGAGAGCCCCACCTCTTGATATGAGAGCCCCACCTCTTGATATGAGAGCCCCACCTC	421
QY	548	TCGGAATATGAGAGCCCCACCTGAGAGAAATGAAGGCGCGCTGCGAG	593
DB	422	TCGGAATATGAGAGCCCCACCTGAGAGAAATGAAGGCGCGCTTCCGG	467
RESULT 5			
AAS20601			
ID	AAS20601	standard; cDNA; 1413 BP.	
AC	AAS20601;		
DT	09-APR-2002	(first entry)	
DE	DNA encoding bovine perinuclear theca 32 (PT32).		
KM	Testicular WM domain binding protein; hMWBP; perinuclear theca 32;		
KM	PT32; contraceptive; fertility; oocyte activation; vaccine;		
KM	globoscopy; spermogenesis; spermatozoa; tyrosine kinase; c-Yes;		
KM	immunoreceptive; bovine; gene; ss.		
OS	Bos sp.		
FX			
FX	Key	Location/Qualifiers	
FT	primer_bind	30..50	
FT		/*tag= a	
FT		/note= "primer binding site for cDNA isolation. The	
FT		sequence differs from that of the forward primer	
FT		given in AAS20603"	
FT	CDS	36..977	
FT		/*tag= b	
FT		/product= "PT32"	
FT		/note= "Perinuclear theca 32"	
FT	protein_bind	978..1001	
FT		/*tag= c	
FT		/note= "primer binding site for cDNA isolation. The	
FT		sequence differs from that of the reverse primer	
FT		given in AAS20604"	
FX			
FX	PN	MO200190185-A2.	
FX	PD	29-NOV-2001.	
FX	PF	25-MAY-2001; 2001MO-CA00738.	
FX	PR	25-MAY-2000; 2000CA-2307128.	
FX	PR	25-MAY-2000; 2000US-206979P.	
FX	PA	(TOOH) UNIV QUEBENS KINGSTON.	
FX	PA	(UYOR-) UNIV OREGON HEALTH SCI.	
FX	PI	Oklo R, Sutovsky P;	
FX	DR	WPI; 2002-097644/13.	
FX	DR	P-PSDB; AAU74604.	
FX	PT	Isolated perinuclear theca 32 polypeptide that interacts with activated	
FX	PT	tyrosine kinase c-Yes, for enhancing fertility, treating/diagnosing	
FX	PT	diminished fertility and abnormal spermiogenesis and for providing	
FX	PT	contraception -	

PS Claim 10; Fig 2A-B; 103pp; English.

CC The invention describes an isolated perinuclear theca 32 (P132)
CC polypeptide (I) which interacts with tyrosine kinase C-Yes. (I) is
CC useful for: enhancing fertility in a mammal; treating globozoospermia, by
CC expressing (I) in spermatozoa; inhibiting fertilisation, by introducing
CC (I) or its antigenic fragment into a mammal to elicit an immune
CC response; enhancing the ability of round spermatozoa to activate oocytes;
CC treating or diagnosing diminished fertility and abnormal spermiogenesis;
CC in providing contraception; identifying contraceptive and
CC fertility-enhancing agents. The polynucleotide is useful for producing
CC (I) by recombinant techniques, as vaccine, as diagnostic reagents, and
CC for chromosome identification. An antibody against (I) is useful in
CC immunological assays. In immunoneutralisative methods, to identify cells
CC expressing (I), and to purify (I) by affinity chromatography. A
CC transgenic animal is useful as an animal model for studying human
CC fertility and reproductive biology, and for screening compounds to
CC identify modulators of oocyte activation. The use of (I) prevents the
CC entry of components which are detrimental to embryonic development into
CC the oocyte during oocyte activation with crude sperm extract and avoids
CC the propagation of viruses such as HIV (human immunodeficiency virus) and
CC SIV (simian immunodeficiency virus) carried in the sperm. This sequence
CC encodes the bovine perinuclear theca 32 (P132), described in the method
CC of the invention.

SQ Sequence 1413 BP; 377 A; 363 C; 369 G; 304 T; 0 other;

Query Match	43.5%	Score	435.4	DB	24	Length	1413
Best Local Similarity	74.7%	Pred. NO.	4.8e-120				
Matches	609	Conservative	0	Mismatches	181	Indels	25
						Gaps	4

Oy	1	TTGCCATTGTATCTGATATGACGAACCTCATCTGTGGAAACAACGATATTTGCTGCAAACTTC	60
Db	258	ATGCCGTTTGGCTTATATGATGATGATCTGCACCAATTGAACAACAATTTTTCGCCCAACTTC	317
Oy	61	ATTAAAGGAACCTATTCAAGCAGCTCCATATGCTGCTGGGAAGACAACACTTACTTTTAA	120
Db	318	ATTAAAGGAAACCAATTCAGCAGACTCCAGGTGTGCTGGGAAGACAACACTGTTTAAAG	377
Oy	121	TTAGCTTCAGAAATGGAATGCGCATTTGACCTGTAATGCTGTAATGCTGCTCT	180
Db	378	TTATCTTCAGGAAGAGAGGTGCATTCGATTTGCCCACTATGCTTAAACCTGCTCT	437
Oy	181	GCTGTGCGCCGAGATTTCCACTTGAACCTTAAATGACTGCTTCACTTATGGAAAT	240
Db	438	GCTGTGCGCAAGGAATTCACCTTGAAGTGTAAATTACTGTGTTGACACTTCAGAGCTG	497
Oy	241	TATGTAAATTACTGGGGAAAGGAAT---ATGTCACTCCACAGATGCTTGTTCAG-----	292
Db	498	TACATAAATTACTGTCCAGAGGGCTCAAGTGTCTCTCCACACACACTTGTCCACATAT	557
Oy	293	----TTATGTCTATGAGGCGCCCACTGTGAGATATGAGGCCCACTCCGATATACGA	348
Db	558	CCAAATGTGATCTATGGAACCCCAACAACAGGATATACAGTCCAAACAGGGGAATATGGA	617
Oy	349	GCCCCACTGCAGGATATGAGGCCCAACCTGTAGGAATGAAGGCCCGCTGTGGGATAC	408
Db	618	ACTCCACAAGAGGATATGAGGCCCAACAGGGGGAATGAGGCCCACTATGGAATAT	677
Oy	409	AGAGCTCACTGTGCGATATGAGGCCCAACCTCTTGGATACGAGGCCCACTGCAAGA	468
Db	678	GGAAGCCCGCTGTGTGGATATGGAATCCCACTGTGGGGAATATGAGATCCCACTGGGGA	737
Oy	469	TATGAGCCCCCACTCTAGGATATGAGGCCCACTCTTGGATATGGAACCCCACTCTC	528
Db	738	TATGAGATGCCCACTGTGGGGAATATGAGGCCCACTGTGGGGAATATGAGATCCCACTGGG	797
Oy	529	GGATATGAGCCCCCACTCTGGAATATGAGGCCCACTGACGAGAAATGAAGCCCGCT	588
Db	798	GGATATGAGCCCCCACTGTGGGGAATATGAGGCCCACTGACGAGATATGAGGCCCAACA	857
Oy	589	GCGGATACAGAGCTCACTGTGGAATCAGAGCCAGGCTTCAGGAATCTACAGACGCC	648

Db 858 GCTGGAATGAAAGCCCTACCCCCGCAATGAAAGCTCATCTGCTGGAAATAACAGCTGCC 917

649 -----CAGGCTCTGTAAGCAGGCTTCTCTTCCCTGCTGCTCTCTTTTCAG 696

Db 918 TCTCAGACATCTATGACAGCTCAGCAGAGACTTCTCTTCCCACTACCTCATCTTTTAA 977

697 GTCCATCTTAACTTTCTTAAGATGTAACCTTGAAGACTCACCAGCAAGAGTACCTT 756

Db 978 GTCATTATTCACACCTTTCAGAGTTAAACCTTGAAGACTCACCAGCAAGAG-GGACACCT 1038

697 GTCCATCTTAACTTTCTTAAGATGTAACCTTGAAGACTCACCAGCAAGAGTACCTT 756

Db 757 AAAATTGAAGTCAGATTAAGGAGGACGACTGAGCT 791

1037 AAAACTGAGTTCACGTTAAGAGAGAGCCAGGT 1071

RESULT 6

ID ABA67850 standard; DNA; 436 BP.

AC ABA67850

DT 01-FEB-2002 (first entry)
 YY

DB Human foetal liver single exon nucleic acid probe #16155.

Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX
US
HOMO VAPLENA:

MOZ00131211-AZ
FN
XX

03-AUG-2001
XX

[illegible]

PR 26-MAY-2000; 2000US-0207456.

PR 03-AUG-2000; 2000US-0632366.

PR 27-SEP-2000; 2000US-0236359.

XX

2 XX

XX
XX
WST 2003 403447/50

XX

DE

PT analyzing gene expression in hu
xy

PS Claim 4; SEQ ID NO 16155; 639pp + sequence

measuring human den

CC measuring and displaying gene expression in samples derived from
CC liver: the single exon nuclear acid probes may be used for pred

CC probe of the invention.
CC record driver. The present sequence is a single ex

printed specification, but was obtained in electronic format directly

XX

Best Local Similarity 99.8%; Pred. No. 5.1e-120;

.....

[illegible]

QY 353 CACCTGAGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGATACAGAG 412
 DB 61 CACCTGAGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGATACAGAG 120
 QY 413 CCTGACCTGTGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGATACAGAG 472
 DB 121 CCTGACCTGTGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGATACAGAG 180
 QY 473 GAGCCCAACCTGTAGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGAT 532
 DB 181 GAGCCCAACCTGTAGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGAT 240
 QY 533 ATGAGAGCCCAACCTGTGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGG 592
 DB 241 ATGAGAGCCCAACCTGTGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGG 300
 QY 593 GATACAGAGCCCAACCTGTGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGG 652
 DB 301 GATACAGAGCCCAACCTGTGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGG 360
 QY 653 CTCCTGAAAACGAGGCTTCTCTCCCTCTGAGGCTCTCTCTGAGGTCATTCTTAACCTT 712
 DB 361 CTCCTGAAAACGAGGCTTCTCTCCCTCTGAGGCTCTCTCTGAGGTCATTCTTAACCTT 420
 QY 713 CTAAAGATGTAAACCTT 728
 DB 421 CTAAAGATGTAAACCTT 436

RESULT 7
 AAK42003
 ID AAK42003 standard; DNA; 436 BP.

AAK42003;

06-NOV-2001 (first entry)

Human bone marrow expressed single exon probe SEQ ID NO: 16560.

Human; bone marrow expressed exon; gene expression analysis; probe;
 microarray; cancer; leukemia; lymphoma; myeloma; ss.

Homo sapiens.

WO200157276-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00668.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488900/53.

Human genome-derived single exon nucleic acid probes useful for
 analyzing gene expression in human bone marrow -
 Example 4; SEQ ID NO: 16560; 658bp + Sequence Listing; English.
 The present invention provides a number of single exon nucleic acid
 probes which are derived from genomic sequences expressed in the human
 bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is one of
 CC the probes of the invention.

Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;

Query Match 43.4%; Score 434.4; DB 22; Length 436;
 Best Local Similarity 99.8%; Pred. No. 5.1e-120;
 Matches 435; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 293 TTATTGTCTATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGATACAGAG 352
 DB 1 TTATTGTCTATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGATACAGAG 60
 QY 353 CACCTGAGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGATACAGAG 412
 DB 61 CACCTGAGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGATACAGAG 120
 QY 413 CCTGACCTGTGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGATACAGAG 472
 DB 121 CCTGACCTGTGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGATACAGAG 180
 QY 473 GAGCCCAACCTGTAGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGAT 532
 DB 181 GAGCCCAACCTGTAGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGGAT 240
 QY 533 ATGAGAGCCCAACCTGTGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGG 592
 DB 241 ATGAGAGCCCAACCTGTGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGG 300
 QY 593 GATACAGAGCCCAACCTGTGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGG 652
 DB 301 GATACAGAGCCCAACCTGTGATATGAGAGCCCAACCTGTAGAAATGAAAGCCCGCTGTGG 360
 QY 653 CTCCTGAAAACGAGGCTTCTCTCCCTCTGAGGCTCTCTCTGAGGTCATTCTTAACCTT 712
 DB 361 CTCCTGAAAACGAGGCTTCTCTCCCTCTGAGGCTCTCTCTGAGGTCATTCTTAACCTT 420
 QY 713 CTAAAGATGTAAACCTT 728
 DB 421 CTAAAGATGTAAACCTT 436

RESULT 8

AAI48070
 ID AAI48070 standard; DNA; 436 BP.

AAI48070;

17-OCT-2001 (first entry)

Probe #16756 used to measure gene expression in human placenta sample.

Probe; microarray; human; placenta; antenatal diagnosis;
 genetic disorder; ss.

Homo sapiens.

WO200157272-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00663.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488897/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 16756; 654bp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 43.4%; Score 434.4; DB 22; Length 436;
 Query Match Best Local Similarity 99.8%; Pred. No. 5.1e-120;
 Matches 435; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 293 TTATGCTCTATGAGGCCCCACCTGAGATATGAGCCCCACCTCCGAGATACGAGCCC 352
 DB 1 TTATGCTCTATGAGGCCCCACCTGAGATATGAGCCCCACCTCCGAGATACGAGCCC 60
 QY 353 CACCTGAGATATGAGGCCCCACCTGAGATATGAGGCCCCACCTCCGAGATACGAG 412
 DB 61 CACCTGAGATATGAGGCCCCACCTGAGATATGAGGCCCCACCTCCGAGATACGAG 120
 QY 413 CCTGACCTGAGATATGAGGCCCCACCTGAGATATGAGGCCCCACCTCCGAGATACGAG 472
 DB 121 CCTGACCTGAGATATGAGGCCCCACCTGAGATATGAGGCCCCACCTCCGAGATACGAG 180
 QY 473 GAGCCCCACCTGAGATATGAGGCCCCACCTGAGATATGAGGCCCCACCTCCGAGAT 532
 DB 181 GAGCCCCACCTGAGATATGAGGCCCCACCTGAGATATGAGGCCCCACCTCCGAGAT 240
 QY 533 ATGAGGCCCCACCTGAGATATGAGGCCCCACCTGAGATATGAGGCCCCACCTCCGAG 592
 DB 241 ATGAGGCCCCACCTGAGATATGAGGCCCCACCTGAGATATGAGGCCCCACCTCCGAG 300
 QY 593 GATACAGGCTCCTGAGATATGAGGCCCCACCTGAGATATGAGGCCCCACCTCCGAG 652
 DB 301 GATACAGGCTCCTGAGATATGAGGCCCCACCTGAGATATGAGGCCCCACCTCCGAG 360
 QY 653 CTCTGAAAACGAGGCTTCTTCCCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 712
 DB 361 CTCTGAAAACGAGGCTTCTTCCCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
 QY 713 CTAAAGATGTAAACCTT 728
 DB 421 CTAAAGATGTAAACCTT 436
 RESULT 9
 ABS16034
 ID ABS16034 standard; DNA; 436 BP.
 XX
 AC ABS16034;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DB Human genome-derived single exon probe ORF from lung SEQ ID No 16025.
 XX
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhage;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Kartagener syndrome;
 KW primary ciliary dyskinesia; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;

KW hyaline membrane disease; open reading frame; ORF.
 XX
 OS Homo sapiens.
 XX
 FN W0200186003-A2.
 PD 15-NOV-2001.
 XX
 PP 30-JAN-2001; 2001WO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLB-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2002-114183/15.
 DR
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX
 PS Claim 4; SEQ ID No 16025; 634bp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridize at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridization of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single probe
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridization to a single exon
 CC microarray having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probe/open reading frames (ORFs). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemorrhage, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Kartagener syndrome, fibrocystic
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 43.4%; Score 434.4; DB 24; Length 436;
 Query Match

Best Local Similarity 99.8%; Pred. No. 5,1e-120;
Matches 435; Conservative 0; Mismatches 1; Indels 0; Gaps 0

OY	293	TTATATGCTCATAGGGGGCCCCACCTGAGAGATATAGAGGCCCACTCCCGGATACGAGGCC	355
Db	1	TTATATGCTCATAGAGCCCCACCTGAGAGATATAGAGGCCCACTCCCGGATACGAGGCC	60
OY	353	CACCTGCAGAGATATGAGGCCCAACCGTAGAAATGAAAGCCCGCTGTGGATACGAG	412
Db	61	CACCTGCAGAGATATGAGGCCCAACCGTAGAAATGAAAGCCCGCTGTGGATACGAG	120
OY	413	CCTCACTGTGTGGATATGAGGCCCACTCTTGGATACGAGGCCCACTGCAGATATG	472
Db	121	CCTCACTGTGTGGATATGAGGCCCACTCTTGGATACGAGGCCCACTGCAGATATG	180
OY	473	GAGCCCCACCTTATGAGATATGAGGCCCACTCTTGGATATGAGGCCCACTGCAGAT	532
Db	181	GAGCCCCACCTTATGAGATATGAGGCCCACTCTTGGATATGAGGCCCACTGCAGAT	240
OY	533	ATGAGAGCCCACTCTCGGATATGAGGCCCACTTCAGAAATGAAAGCCCGCTGCG	592
Db	241	ATGAGAGCCCACTCTCGGATATGAGGCCCACTTCAGAAATGAAAGCCCGCTGCG	300
OY	593	GATACAGAGCCTCACTGTGATACGAGGCCCAAGCCTCAGAGATCTTACAGAGCCOAG	652
Db	301	GATACAGAGCCTCACTGTGATACGAGGCCCAAGCCTCAGAGATCTTACAGAGCCOAG	360
OY	653	CTCTCGAAAAACGAGGCTTCTCTTCCCTCTGCGCCCTCTCTCAGAGTCATCTTAACCTT	712
Db	361	CTCTCGAAAAACGAGGCTTCTCTTCCCTCTGCGCCCTCTCTCAGAGTCATCTTAACCTT	420
OY	713	CTAAGATGTAAACCTT 728	
Db	421	CTAAGATGTAAACCTT 436	
RESULT 10			
ID	AA104882	AA104882 standard; DNA; 7099 BP.	
AC	AA104882;		
DT	21-NOV-2001 (first entry)		
XX	Human reproductive system related antigen DNA SEQ ID NO: 7570.		
XX	Human; reproductive system related antigen; reproductive system disorder;		
KW	cancer; gene therapy; ds.		
XX	Homo sapiens.		
XX	WO200155320-A2.		
XX	02-AUG-2001.		
PD	17-JAN-2001; 2001MO-US01339.		
XX			
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		

QY 410 GAGCTCACTGTGGATATGAGCCCACTTCTTGATACGAGCCCACTTGAGGAT 469
 DB 6843 GAGCTCACTGTGGATATGAGCCCACTTCTTGATACGAGCCCACTTGAGGAT 6902
 QY 470 ATGAGGCCCACTTCTTGATATGAGCCCACTTCTTGATATGAGCCCACTTCTTG 529
 DB 6903 ATGAGGCCCACTTCTTGATATGAGCCCACTTCTTGATATGAGCCCACTTCTTG 6962
 QY 530 GATATGAGCCCACTTCTTGATATGAGCCCACTTCTTGATATGAGCCCACTTCTTG 589
 DB 6963 GATATGAGCCCACTTCTTGATATGAGCCCACTTCTTGATATGAGCCCACTTCTTG 7022
 QY 530 CGGATATGAGCCCACTTCTTGATATGAGCCCACTTCTTGATATGAGCCCACTTCTTG 649
 DB 7023 CGGATATGAGCCCACTTCTTGATATGAGCCCACTTCTTGATATGAGCCCACTTCTTG 7082
 QY 650 AGGCTCTGAAAACGAG 666
 DB 7083 AGGCTCTGAAAACGAG 7099

RESULT 12

AA570582
 ID AA570582 standard; cDNA; 894 BP.
 AC AA570582;
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #6386.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

FN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG06395.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

PS Claim 1; SEQ ID NO 6386; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABA5197-ABA54564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIR0
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 894 BP; 204 A; 221 C; 213 G; 256 T; 0 other;

QY Query Match 17.3%; Score 173.6; DB 23; Length 894;
 DB Best Local Similarity 86.8%; Pred. No. 1.6e-41;
 DB Matches 191; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 448 TACGAGCCCACTTCTTGATATGAGCCCACTTCTTGATATGAGCCCACTTCTT 507
 DB 165 TATGAGCCCACTTCTTGATATGAGCCCACTTCTTGATATGAGCCCACTTCTT 224
 QY 508 GATATGAGCCCACTTCTTGATATGAGCCCACTTCTTGATATGAGCCCACTTCTT 567
 DB 225 GATATGAGCCCACTTCTTGATATGAGCCCACTTCTTGATATGAGCCCACTTCTT 284
 QY 568 GCAGAAATGAGCCCACTTCTTGATATGAGCCCACTTCTTGATATGAGCCCACTTCTT 627
 DB 285 GCAGAAATGAGCCCACTTCTTGATATGAGCCCACTTCTTGATATGAGCCCACTTCTT 344
 QY 628 CCTCAGAAATCTTACGAGCCCACTTCTTGAAACGAG 667
 DB 345 CCTCAGAAATCTTACGAGCCCACTTCTTGAAACGAG 384

RESULT 13

ABA55246
 ID ABA55246 standard; DNA; 471 BP.

AC ABA55246;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #3551.

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

FN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human foetal liver -

XX Claim 1; SEQ ID NO 3551; 639bp + sequence listing; English.
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WFO at ftp.wfo.int/pub/published_pct_sequences.
XX
SQ Sequence 471 BP; 120 A; 112 C; 82 G; 157 T; 0 other;
Query Match 13.0%; Score 130.4; DB 22; Length 471;
Best Local Similarity 99.2%; Pred. No. 1.1e-28;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 290 CAGTTATGTCTATGAGGGCCCCACCTGCAGGATATGAGCCCCACCTCCCGATACGAG 349
DB 340 CAGTTATGTCTATGAGAGCCCCACCTGCAGGATATGAGCCCCACCTCCCGATACGAG 399
QY 350 CCCCACTGCAGGATATGAGAGCCCCACCTGCAGGATATGAGAGGCGCGCTGTGGATACA 409
DB 400 CCCCACTGCAGGATATGAGAGCCCCACCTGCAGGATATGAGAGGCGCGCTGTGGATACA 459
QY 410 GAGCCTCACCTG 421
DB 460 GAGCCTCACCTG 471
RESULT 14
AAK28957
ID AAK28957 standard; DNA; 471 BP.
XX
AC AAK28957;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 3514.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 3514; 658bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.

XX
SQ Sequence 471 BP; 120 A; 112 C; 82 G; 157 T; 0 other;
Query Match 13.0%; Score 130.4; DB 22; Length 471;
Best Local Similarity 99.2%; Pred. No. 1.1e-28;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 290 CAGTTATGTCTATGAGGGCCCCACCTGCAGGATATGAGCCCCACCTCCCGATACGAG 349
DB 340 CAGTTATGTCTATGAGAGCCCCACCTGCAGGATATGAGCCCCACCTCCCGATACGAG 399
QY 350 CCCCACTGCAGGATATGAGAGCCCCACCTGCAGGATATGAGAGGCGCGCTGTGGATACA 409
DB 400 CCCCACTGCAGGATATGAGAGCCCCACCTGCAGGATATGAGAGGCGCGCTGTGGATACA 459
QY 410 GAGCCTCACCTG 421
DB 460 GAGCCTCACCTG 471
RESULT 15
AAI34906
ID AAI34906 standard; DNA; 471 BP.
XX
AC AAI34906;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #3592 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488997/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 3592; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 471 BP; 120 A; 112 C; 82 G; 157 T; 0 other;
Query Match 13.0%; Score 130.4; DB 22; Length 471;
Best Local Similarity 99.2%; Pred. No. 1.1e-28;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 290 CAGTTATGTCTATGAGGGCCCCACCTGCAGGATATGAGCCCCACCTCCCGATACGAG 349

Db	340	CAATTATTGTCTATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAG	399
Qy	350	CCCCACCTGCGAGATATGAGCCCAACCCGTAGGAATGAAGCCCGCTGTGGATACA	409
Db	400	CCCCACCTGCGAGATATGAGCCCAACCCGTAGGAATGAAGCCCGCTGTGGATACA	459
Qy	410	GAGCCTGACCTG	421
Db	460	GAGCCTGACCTG	471

Search completed: December 16, 2003, 11:00:50
Job time : 270.133 secs

TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT. . .
FILE REFERENCE: XX/P04470U50
CURRENT APPLICATION NUMBER: US/08/841,349B
CURRENT FILING DATE: 1997-04-30
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 981
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: For all n's in this sequence, n=(a or g or c or t)
US-08-841-349-15

Query Match
Best Local Similarity 53.2%; Score 76; DB 2; Length 981;
Best Local Similarity 53.2%; Pred. No. 9.5e-14;
Matches 160; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

314 CTGACGATATGAGAGCCCACTCCCGGATACGAGAGCCCACTGACGATATGAGGCC 373
DB CTCCAGGCGCATGTAGCCAGCTCTCCAGGCGCATGTAGCCAGCTCTCCAGGCGCATGTAGCC 889
QY 374 AACCCGTAGGAATGAAGCCCGCTGTGGATACAGAGCTCACTGTGCGATATGAG 433
DB 888 GCTTCCAGGCGCATGTAGCCAGCTCTCCAGGCGCATGTAGCCAGCTCTCCAGGCGCATGTAG 829
QY 434 CCCACCTCTTGAATACGAGAGCCCACTGACGATATGAGAGCCCACTGTAGATATG 493
DB 828 CCAAGCTCTCCAGGCGCATGTAGCCAGCTCTCCAGGCGCATGTAGCCAGCTCTCCAGGCGCAT 769
QY 494 GACCCCACTCTTGAATATGAGAGCCCACTCTCGATATGAGAGCCCACTCTCGAT 553
DB 768 TAGGCGATATGAGAGCATGTAGGCGCTCTCCAGGCGCATGTAGCCAGCTCTCCAGGCC 709
QY 554 ATGAGAGCCCACTGACGATATGAGAGCCCGCTGTGGATACAGAGCTCACTGTGCTG 613
DB 708 ATGTGGGCGAGCTCTCCAGGCGCATGTAGGCGAGCTCTCCAGGCGCATATTTGGAGCTCTCCAG 649
QY 614 G 614
DB 648 G 648

RESULT 3
US-08-182-175A-56
Sequence 56, Application US/08182175A
Patent No. 5559223
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSER: E. I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,175A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:

NAME: Linda Axamechy, Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DH5 alpha
IMMEDIATE SOURCE:
CLONE: 2-9
FEATURE:
NAME/KEY: CDS
LOCATION: 2..235
OTHER INFORMATION: /function= "synthetic storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "gsp"
OTHER INFORMATION: /standard_name= "7.7.7.7.7.8.9.8.9.5"
US-08-182-175A-56

Query Match
Best Local Similarity 53.2%; Score 54.4; DB 1; Length 243;
Best Local Similarity 53.2%; Pred. No. 2.4e-07;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

302 ATGGGGCCCCACCTGAGATATGAGAGCCCACTCCCGGATACGAGAGCCCACTGACG 361
DB 2 ATGAGAGAGAACTGAGAGCGATGAGAGAGAGCTGAAGCGCATGAGAGAGAGCTGAAG 61
QY 362 GATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGTGGATACAGAGCTCACTG 421
DB 62 GCGATGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAGCGCATGAGAGAGAGCTGAAG 121
QY 422 TCGATATGAGAGCCCACTCTTGAATACGAGAGCCCACTGACGATATGAGAGCCCACT 481
DB 122 AAGCGATGAGAGAGAAAGCTTAAGAGAGATGAGAGAGAAAGCTTAAGATGAGAGAGAA 181
QY 482 CTCTAGATATGAGAGCCCACTCTTGAATATGAGAA 517
DB 182 CTCAAAAAGATGAGAGAAAGCTTAATGATGAGAA 217

RESULT 4
US-08-474-633A-74
Sequence 74, Application US/08474633A
Patent No. 5773691
GENERAL INFORMATION:
APPLICANT: COMPANY
APPLICANT: COMPANY
TITLE OF INVENTION: CHIMERIC GENES AND
TITLE OF INVENTION: METHODS FOR INCREASING
TITLE OF INVENTION: INCREASING THE LYSINE
TITLE OF INVENTION: AND THREONINE CONTENT
TITLE OF INVENTION: OF THE SEEDS OF PLANTS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSER: E. I. DU PONT DE NEMOURS
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE

```

1 OPERATING SYSTEM: PC-DOS/MS-DOS
2 SOFTWARE: MICROSOFT WORD VERSION 2.00C
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: US/08/474,633A
5 FILING DATE:
6 CLASSIFICATION: 800
7 ATTORNEY/AGENT INFORMATION:
8 NAME: BARBARA C. SIEGEL
9 REGISTRATION NUMBER: 30,684
10 REFERENCE/DOCKET NUMBER: BB-1037-C
11 TELECOMMUNICATION INFORMATION:
12 TELEPHONE: 302-992-4931
13 TELEFAX: 302-773-0164
14 TELEX: 835420
15 INFORMATION FOR SEQ ID NO: 74:
16 SEQUENCE CHARACTERISTICS:
17 LENGTH: 243 base pairs
18 TYPE: nucleic acid
19 STRANDEDNESS: double
20 TOPOLOGY: linear
21 MOLECULE TYPE: DNA (genomic)
22 ORIGINAL SOURCE:
23 STRAIN: E. coli
24 CELL TYPE: DH5 alpha
25 IMMEDIATE SOURCE:
26 CLONE: 2-9
27 FEATURE:
28 NAME/KEY: CDS
29 LOCATION: 2..235
30 OTHER INFORMATION: /function= "synthetic"
31 OTHER INFORMATION: storage protein
32 OTHER INFORMATION: /product= "protein"
33 OTHER INFORMATION: /gene= "gsp"
34 OTHER INFORMATION: /standard name=
35 OTHER INFORMATION: "7.7.7.7.7.8.9.8.9.5T"
36 US-08-474-633A-74

```

```

Query Match Similarity      5.4%; Score 54.4; DB 1; Length 243;
Best Local Similarity       53.2%; Pred No. 2.4e-07;
Matches 115; Conservative   0; Mismatches 101; Indels    0; Gaps    0

Dy      302 ATGGGCCCCCACCCTGCAGATATGTGAGCCCCCACTCCCGATACGAGGCCCACTGCAG 361
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       2  ATGAGGAGAAGCTGAAGCGCATGTAGAGGAAGACTGAAGCGCATGAGGAGAAGCTGAAG 61
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy      362 GATATGAGGCCCAACCCCTAGAAATTGAAGGCCCCCTGTGGGATTCAGAGCTTCACTTG 421
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
        62 GCGATGAGGAGAAAGCTAAGGCGCATGTAGAGGAAGCTGAAAGCCATGAGGAAGCTG 121
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy      422 TGCATATGTAGAGCCCCCACCTCTTTGATATCGAGGCCCACTTGCAGATATGTGAGCCCCCAG 481
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
        122 AAGCGATGTAGAGAAAAGCTTAAGTAGATGAGAAAAAGCTGAATTGATGTAGAGAGAAA 181
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy      482 CTCTAGATATGTAGAGCCCCCACCTCTTGTAGATATGAAA 517
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
        182 CTCAAAAAGATGAGAGAAAAGCTTAATGATGTAGAAA 217
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
US-08-823-771-74
; Sequence 74, Application US/08823771
; Patent No. 6459019
GENERAL INFORMATION:
APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY
TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR INCREASING THE LYSINE INCREASING THE LYSINE AND THREONINE CONTENT
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY

```

```

1 STREET: 1007 MARKET STREET
2 CITY: WILMINGTON
3 STATE: DELAWARE
4 COUNTRY: U.S.A.
5 ZIP: 19898
6
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: FLOPPY DISK
9 COMPUTER: IBM PC COMPATIBLE
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11 SOFTWARE: MICROSOFT WORD VERSION 2.0C
12
13 CURRENT APPLICATION DATA:
14 APPLICATION NUMBER: US/08/823,771
15 FILING DATE: 24-Mar-1997
16 CLASSIFICATION: <Unknown>
17
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: 08/474,633
20 FILING DATE: <Unknown>
21 ATTORNEY/AGENT INFORMATION:
22 NAME: BARBARA C. SIEGEL
23 REGISTRATION NUMBER: 30,684
24 REFERENCE/DOCKET NUMBER: BB-1037-C
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE: 302-992-4931
27 TELEFAX: 302-773-0164
28 TELEX: 835420
29
30 INFORMATION FOR SEQ ID NO: 74:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 243 base pairs
33 TYPE: nucleic acid
34 STRANDEDNESS: double
35 TOPOLOGY: linear
36
37 MOLECULE TYPE: DNA (genomic)
38 ORIGINAL SOURCE:
39 STRAIN: E. coli
40 CELL TYPE: DHS alpha
41 IMMEDIATE SOURCE:
42 CLONE: 2-9
43
44 FEATURE:
45 NAME/KEY: CDS
46 LOCATION: 2..235
47 OTHER INFORMATION: /function= "synthetic
48 storage protein
49 /product= "protein"
50 /gene= "sep"
51 /standard_name=
52 "7.7.7.7.7.8.9.8.9.5"
53
54 SEQUENCE DESCRIPTION: SEQ ID NO: 74:
55
56 US-08-823-771-74
57
58 Query Match 5.4%; Score 54.4; DB 4; Length 243;
59 Best Local Similarity 53.2%; Pred. No. 2.4e-07;
60 Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
9
```

RESULT 6
PCT-US92-06412-56
; Sequence 56, Application PC/TUS9206412

GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing R
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E.I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06412
FILING DATE: 19920807
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axamechy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DHS alpha
IMMEDIATE SOURCE:
CLONE: 2-9
FEATURE:
NAME/KEY: CDS
LOCATION: 2..235
OTHER INFORMATION: /function= "synthetic storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "gsp"
OTHER INFORMATION: /standard_name= "7.7.7.7.8.9.8.9.5"
PCT-US92-06412-56

Query Match 5.4%; Score 54.4; DB 5; Length 243;
Best Local Similarity 53.2%; Pred. No. 2.4e-07;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 302 ATGGGGCCCCACCTGAGATATGAGCCCACTCCGAGATACGAGCCCACTGACG 361
DB 2 ATGGAGGAGAACTGAGAGCGATGAGAGAGAGCTGAGAGCGATGAGAGAGCTGAG 61
QY 362 GATATGAGCCCAACCCGATGAGAAATGAAAGCCCGCTGTGATACAGAGCTCACTG 421
DB 62 GCGATGAGAGAGAGCTGAGAGCGATGAGAGAGAGCTGAGAGCGATGAGAGAGCTG 121
QY 422 TGCATATGAGAGCCCACTCTTGTGATACGAGCCCACTGAGATATGAGAGCCCAAC 481
DB 122 AAGGAGATGAGAGAGAAAGCTTAAAGAGATGAGAAAGAGCTGAAATGATGAGAGAGAA 181
QY 482 CTCTAGATATGAGAGCCCACTCTTGTGATATGAA 517
DB 182 CTCAAAAGATGAGAGAGAAAGCTTAAATGATGAA 217

RESULT 7
US-08-917-320-18
Sequence 18, Application US/08917320
Patent No. 5824508
GENERAL INFORMATION:
APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
TITLE OF INVENTION: No. 5824508 Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/917,320
FILING DATE: 25-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Luann Coert
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5163
TELEFAX: 415-857-0663
TELEX: 380816 CooleyPA
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1014..3734
US-08-917-320-18

Query Match 5.4%; Score 53.6; DB 1; Length 3833;
Best Local Similarity 50.4%; Pred. No. 2.3e-06;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 307 GCCCACCCTGAGATATGAGCCCACTCCGAGATACGAGCCCACTGACGATAT 366
DB 2578 GCCCACCCTGAGATATGAGCCCACTCCGAGATACGAGCCCACTGACGATAT 2637
QY 367 GAGCCCAACCCGATGAGAAATGAAAGCCCGCTGTGATACAGAGCTCACTGTGGA 426
DB 2638 CTACCCCAACCCCAATGAGCAAGCCCACTTGGGAAAAACAAGTCTTACACAG 2697
QY 427 TATGAGCCCACTCTTGTGATATGAGCCCACTGAGATATGAGAGCCCACTCTA 486
DB 2698 TACTTACCCCAACCCCAATGAGCAAGCCCACTTGGGAAAAACAAGCCCACTTGA 2757
QY 487 GATATGAGCCCACTCTTGTGATATGAGCCCACTCTGATATGAGAGCCCACT 546
DB 2758 CAGTGAATACCCCAACCCCAATGAGCAAGCCCACTTGGGAAAAACAAGCCCACT 2817
QY 547 CTGGAATATGAGAGCCCAAC 566
DB 2818 CAGCAGTGAATACCCCAAC 2837

RESULT 8
PCT-US95-04611A-18
Sequence 18, Application PC/TUS9504611A
GENERAL INFORMATION:
APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
TITLE OF INVENTION: Non Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04611A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Luann Cseert
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5163
TELEFAX: 415-857-0663
TELEX: 380816 CooleyPA
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1014..3734
PCT-US95-04611A-18

Query Match
Best Local Similarity 5.4%; Score 53.6; DB 5; Length 3833;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

307 GCCCCACCTGACGATATGAGCCCCCACTCCCGATACGAGCCCCCACTGACGATAT 366
DB 2578 GCCCCACCTGACGATATGAGCCCCCACTCCCGATACGAGCCCCCACTGACGATAT 2637
367 GAGGCCCAACCCGTAAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCGA 426
DB 2638 CTACCCCAACCCCAATGCGACACGAGCCCACTTGGGAAAAACAAGTCTTACTGACGAG 2697
427 TATGAGCCCCCACTCTTGATATGAGAGCCCACTGCGATATGAGAGCCCACTCTTA 486
DB 2698 TGACTACCCCAACCCCAATGCGACACGAGCCCACTTGGGAAAAACAAGTCTTACTGACGAG 2757
487 GATATGAGAGCCCACTCTTGATATGAGAGCCCACTGCGATATGAGAGCCCACT 546
DB 2758 CAGTACCTACCCCAACCCCAATGCGACACGAGCCCACTTGGGAAAAACAAGTCTTACT 2817
547 CTCGATATGAGAGCCCACT 566
DB 2818 CAGCAGTACCTACCCCACT 2837

RESULT 9
US-08-783-774-1
Sequence 1, Application US/08783774
Patent No. 6054130
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
TITLE OF INVENTION: JACKMAN, Winthrop
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/783,774
FILING DATE: 15-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Cortuzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7682-037
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5931 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1014..3734
OTHER INFORMATION:
US-08-783-774-1

Query Match
Best Local Similarity 5.4%; Score 53.6; DB 3; Length 5931;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

307 GCCCCACCTGACGATATGAGCCCCCACTCCCGATACGAGCCCCCACTGACGATAT 366
DB 2578 GCCCCACCTGACGATATGAGCCCCCACTCCCGATACGAGCCCCCACTGACGATAT 2637
367 GAGGCCCAACCCGTAAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCGA 426
DB 2638 CTACCCCAACCCCAATGCGACACGAGCCCACTTGGGAAAAACAAGTCTTACTGACGAG 2697
427 TATGAGCCCCCACTCTTGATATGAGAGCCCACTGCGATATGAGAGCCCACTCTTA 486
DB 2698 TGACTACCCCAACCCCAATGCGACACGAGCCCACTTGGGAAAAACAAGTCTTACTGACGAG 2757
487 GATATGAGAGCCCACTCTTGATATGAGAGCCCACTGCGATATGAGAGCCCACT 546
DB 2758 CAGTACCTACCCCAACCCCAATGCGACACGAGCCCACTTGGGAAAAACAAGTCTTACT 2817
547 CTCGATATGAGAGCCCACT 566
DB 2818 CAGCAGTACCTACCCCACT 2837

RESULT 10

US-09-556-706B-1
; Sequence 1, Application US/09556706B
; Patent No. 6458364
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Jackman, Winthrop
; TITLE OF INVENTION: NON SPLICING VARIANTS OF GP350/220
; FILE REFERENCE: 7682-050-999
; CURRENT APPLICATION NUMBER: US/09/556,706B
; PRIOR APPLICATION NUMBER: 08/783,774
; PRIOR FILING DATE: 1997-01-15
; PRIOR APPLICATION NUMBER: 08/229,291
; PRIOR FILING DATE: 1994-04-18
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 5931
; TYPE: DNA
; ORGANISM: Virus
; FEATURE:
; OTHER INFORMATION: GP350/220
US-09-556-706B-1

Query Match 5.4%; Score 53.6; DB 4; Length 5931;
Best Local Similarity 50.4%; Pred. No. 3e-06;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 307 GCGCCACCTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 366
DB 2578 GCGCCACCTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 2637
QY 367 GGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 426
DB 2638 TGACTACCTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 2697
QY 427 TATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 486
DB 2698 TGACTACCTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 2757
QY 487 GGATATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 546
DB 2758 CAGTATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 2817
QY 547 CTCGATATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 566
DB 2818 CAGCAGTATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 2837

RESULT 11
US-07-865-662F-10
; Sequence 10, Application US/07865662F
; Patent No. 5451670
; GENERAL INFORMATION:
; APPLICANT: Marcia M. Miller
; TITLE OF INVENTION: Restriction Fragment Length
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESS: City of Hope
; STREET: 1500 East Duarte Road
; CITY: Duarte
; STATE: California
; COUNTRY: United States of America
; ZIP: 91010-0269
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3M Double Density 5 1/4" diskette
; OPERATING SYSTEM: MS DOS Version 3.20
; SOFTWARE: Microsoft
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/865,662F
; FILING DATE: 07 April, 1992

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/688,326
FILING DATE: 22 April 1991
APPLICATION NUMBER: 07/588,922
FILING DATE: 27 September 1990
APPLICATION NUMBER: 07/210,405
FILING DATE: 23 June 1988
APPLICATION NUMBER: US 07/130,529
FILING DATE: 9 December 1987
APPLICATION NUMBER: US 07/068,176
FILING DATE: 30 June 1987
ATTORNEY/AGENT INFORMATION:
NAME: Icons, Edward S.
REGISTRATION NUMBER: 16,541
REFERENCE/DOCKET NUMBER: NO. 5451670e
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 783-6040
TELEFAX: (202) 783-6031
TELEX: No. 5451670e
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2188
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: DNA
ORIGINAL SOURCE: Synthetically Prepared
IMMEDIATE SOURCE: Synthetically Prepared
US-07-865-662F-10

Query Match 5.2%; Score 52; DB 1; Length 2188;
Best Local Similarity 47.5%; Pred. No. 5.1e-06;
Matches 154; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 302 ATGGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 361
DB 1315 ATGGAGCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 1374
QY 362 GATATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 421
DB 1375 GAGATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 1434
QY 422 TGCGATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 481
DB 1435 GAGGAGATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 1494
QY 482 CTGAGATATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 541
DB 1495 CATGAGATATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 1554
QY 542 CACTTCTGAGATATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 601
DB 1555 CACTTCTGAGATATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 1614
QY 602 CACTTCTGAGATATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 625
DB 1615 GAGCAGATATGAGCCCACTGAGATATGAGCCCACTCCGATGAGAGCCCACTGAGATAT 1638

RESULT 12
US-08-374-219B-10
; Sequence 10, Application US/08374219B
; Patent No. 6218106
; GENERAL INFORMATION:
; APPLICANT: Miller, Marcia M.
; TITLE OF INVENTION: Restriction Fragment Length Polymorphism Test
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Rothwell, Pigg, Ernst & Kurz
; STREET: 555 Thirteenth Street, N.W.
; CITY: Washington

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: December 16, 2003, 10:43:10 ; Search time 329.43 Seconds

(without alignments)
10099.015 Million cell updates/sec

Title: US-09-864-291-11

Perfect score: 1001

Sequence: 1 atgcattcattcgtgacgac.....attaccctcattgcagtag 1001

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2201672 seqs, 166179959 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	446.2	44.6	467	11	US-09-764-891-1230 Sequence 1230, Ap
2	434.4	43.4	436	9	US-09-864-761-31561 Sequence 31561, Ap
3	375.4	37.5	7099	11	US-09-764-891-7570 Sequence 7570, A
4	130.4	13.0	471	11	US-09-864-761-15029 Sequence 15029, A
5	110.2	11.0	321	11	US-09-764-891-1356 Sequence 1356, Ap
6	110.2	11.0	5769	11	US-09-764-891-7572 Sequence 7572, Ap
7	109	10.9	866	11	US-10-029-386-22839 Sequence 22839, A
8	100	10.0	1040	13	US-10-029-386-26033 Sequence 26033, A
9	93.2	9.3	291	11	US-09-764-891-7571 Sequence 7571, Ap
10	89.4	8.9	1984	9	US-09-864-761-4562 Sequence 4562, Ap
11	88.8	8.9	541	13	US-10-029-386-12333 Sequence 12333, A
12	87.8	8.8	543	9	US-10-029-386-24981 Sequence 24981, A
13	87	8.7	543	9	US-09-864-761-21308 Sequence 21308, A
14	77.2	7.7	432	11	US-09-918-995-33536 Sequence 33536, A
15	69.2	6.9	541	13	US-10-029-386-11265 Sequence 11265, A

C	16	67.8	6.8	511	13	US-10-029-386-24981	Sequence 24981, A
	17	64.8	6.5	593	13	US-10-027-632-290832	Sequence 290832, A
	18	64.8	6.5	593	14	US-10-027-632-290832	Sequence 290832, A
	19	63.2	6.3	409	10	US-09-960-352-6984	Sequence 6984, Ap
	20	63	6.3	509	9	US-09-864-761-25366	Sequence 25366, A
	21	63	6.3	596	9	US-09-864-761-8648	Sequence 8648, Ap
	22	62.2	6.2	689	13	US-10-027-632-148124	Sequence 148124, A
	23	62.2	6.2	689	14	US-10-027-632-148124	Sequence 148124, A
	24	57.4	5.7	2489	11	US-09-298-523B-77	Sequence 72, Appl
	25	57.2	5.7	436	9	US-09-864-761-31561	Sequence 31561, A
	26	57.2	5.7	7099	11	US-09-764-891-7570	Sequence 7570, Ap
	27	54.4	5.4	243	15	US-10-023-066A-74	Sequence 74, Appl
	28	54.2	5.4	1218	9	US-09-864-761-20761	Sequence 20761, A
	29	54.2	5.4	1954	9	US-09-864-761-4001	Sequence 4001, Ap
	30	54.2	5.4	3135	15	US-10-175-523-151	Sequence 151, Ap
	31	54.2	5.4	4758	10	US-09-842-777-9	Sequence 9, Appl
	32	54.2	5.4	5994	13	US-10-205-219-136	Sequence 136, Ap
	33	53.2	5.3	4395	9	US-09-925-301-224	Sequence 224, Ap
	34	53.2	5.3	8412	11	US-09-919-039-223	Sequence 223, Ap
	35	52.8	5.3	2016	10	US-09-938-842A-2004	Sequence 2004, Ap
	36	52.6	5.3	5327	13	US-10-240-453-64	Sequence 64, Appl
	37	52.6	5.3	5327	15	US-10-239-676-62	Sequence 62, Appl
	38	52	5.2	3489	13	US-10-294-804-1	Sequence 1, Appl
	39	50.2	5.0	604	10	US-09-738-973-175	Sequence 175, Ap
	40	50.2	5.0	604	10	US-09-854-133-175	Sequence 175, Ap
	41	50.2	5.0	604	15	US-10-144-642A-175	Sequence 175, Ap
	42	50	5.0	895	9	US-09-864-761-23227	Sequence 23227, A
	43	49.6	5.0	1991	9	US-09-864-761-19477	Sequence 19477, A
	44	49.6	5.0	1959	9	US-09-864-761-2946	Sequence 2946, Ap
	45	49.6	5.0	3463	11	US-09-298-523B-59	Sequence 59, Appl

ALIGNMENTS

RESULT 1

US-09-764-891-1230

1: Sequence 1230, Application US/09764891

2: Publication No. US20030077808A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC006

CURRENT APPLICATION NUMBER: US/09/764,891

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PAM or file wrapper

NUMBER OF SEQ ID NOS: 10231

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1230

LENGTH: 467

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

NAME/KEY: SITE

LOCATION: (442)

OTHER INFORMATION: n equals a,t,g, or c

US-09-764-891-1230

Query Match	44.6%	Score 446.2	DB 11	Length 467
Best Local Similarity	98.5%	Pred. No. 2.6e-117		
Matches 459	Conservative 1	Mismatches 5	Indels 1	Gaps 1
QY	129	CAGAAATGAGATGCAATGAAATTTGGCCAGTTGATGATGAAAGCTGCTGCTGTC	188	
DB	2	CAGAAATGAGATGCAATGAAATTTGGCCAGTTGATGATGAAAGCTGCTGCTGTC	61	
QY	189	CCGAGATTTCCACTTGAACCTTAATGATGCTGTTACCTTATGGAATTTATGTAAT	248	
DB	62	CCGAGATTTCCACTTGAACCTTAATGATGCTGTTACCTTATGGAATTTATGTAAT	121	
QY	249	TACGGGGAAGGAAATGAGCACTTCCAGATGCTTGTCACTTAATGCTATGAGGC	308	
DB	122	TACGGGGAAGGAAATGAGCACTTCCAGATGCTTGTCACTTAATGCTATGAGGC	181	


```

QY 309 CCCACTGAGATATGAGAGCCCACTCCCGGATAGAGAGCCCACTGAGAGATAG 368
   |||||
Db 182 CCCACTGAGATATGAGAGCCCACTCCCGGATAGAGAGCCCACTGAGAGATAG 241
   |||||
QY 369 AGCCCAACCGTATGAGAGAGCCCGCTGTGGATACAGAGCTCACTGTGGAT 428
   |||||
Db 242 AGCCCAACCGTATGAGAGAGCCCGCTGTGGATACAGAGCTCACTGTGGAT 301
   |||||
QY 429 TGAAGCCCACTCTTGTATGAGAGAGCCCACTGAGAGATATGAGAGCCCACTT 488
   |||||
Db 302 TGAAGCCCACTCTTGTATGAGAGAGCCCACTGAGAGATATGAGAGCCCACTT 361
   |||||
QY 489 AATGAGAGCCCACTCTTGTATGAGAGAGCCCACTGAGAGATATGAGAGCCCACT 547
   |||||
Db 362 AATGAGAGCCCACTCTTGTATGAGAGAGCCCACTGAGAGATATGAGAGCCCACT 421
   |||||
QY 548 TCGATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGTGG 593
   |||||
Db 422 TCGATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGTGG 467
   |||||

```

RESULT 2

```

US-09-864-761-31561
; Sequence 31561, Application US/09864761
; Patent No. US20020048763A1
;
GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
;
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117

```

```

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31561
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Homo sapiens
;
FEATURE:
; OTHER INFORMATION: MAP TO 299716.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
; OTHER INFORMATION: EST HUMAN HIT: AM169980.1, EVALUE 8.00e-11
; OTHER INFORMATION: SWISSPROT HIT: P18616, EVALUE 3.00e-13
US-09-864-761-31561

```

```

Query Match 43.4%; Score 434.4; DB 9; Length 436;
Best Local Similarity 99.8%; Pred. No. 2,1e-133;
Matches 435; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 293 TTAATGTCTATGAGAGCCCACTGAGAGATATGAGAGCCCACTCCCGGATACGAGGCC 352
   |||||
Db 1 TTAATGTCTATGAGAGCCCACTGAGAGATATGAGAGCCCACTCCCGGATACGAGGCC 60
   |||||
QY 353 CACTGAGAGATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGTGGATACAGAG 412
   |||||
Db 61 CACTGAGAGATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGTGGATACAGAG 120
   |||||
QY 413 CTTCACTGTGTGATATGAGAGCCCACTTGTGATATGAGAGCCCACTGAGAGATATG 472
   |||||
Db 121 CTTCACTGTGTGATATGAGAGCCCACTTGTGATATGAGAGCCCACTGAGAGATATG 180
   |||||
QY 473 GAGCCCACTCTAGATATGAGAGCCCACTTGTGATATGAGAGCCCACTCTGTGAT 532
   |||||
Db 181 GAGCCCACTCTAGATATGAGAGCCCACTTGTGATATGAGAGCCCACTCTGTGAT 240
   |||||
QY 533 ATGAGAGCCCACTCTGAGATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGTGG 592
   |||||
Db 241 ATGAGAGCCCACTCTGAGATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGTGG 300
   |||||
QY 593 GATACAGAGCTCACTGTGATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGTGG 652
   |||||
Db 301 GATACAGAGCTCACTGTGATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGTGG 360
   |||||
QY 653 CTCCTGAAAGAGAGCTTCTCTCCCTGAGAGATATGAGAGCCCACTCTGAGAGATATG 712
   |||||
Db 361 CTCCTGAAAGAGAGCTTCTCTCCCTGAGAGATATGAGAGCCCACTCTGAGAGATATG 420
   |||||
QY 713 CTAAGATGTAACCTT 728
   |||||
Db 421 CTAAGATGTAACCTT 436
   |||||

```

RESULT 3

```

US-09-764-891-7570
; Sequence 7570, Application US/09764891
; Publication No. US20030077808A1
;
GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; Prior application data removed - consult PAMM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7570
; LENGTH: 7099
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7570

```

```

Query Match 37.5%; Score 375.4; DB 11; Length 7099;
Best Local Similarity 99.7%; Pred. No. 5.1e-113;

```

Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 290 CAGTTATGCTGATGAGGCCCCCAGCTGAGAGATATGAGCCCACTCCGAGATACGAG 349
DB 6723 CAGTTATGCTGATGAGGCCCCCAGCTGAGAGATATGAGCCCACTCCGAGATACGAG 6782

QY 350 CCCCACCTGAGAGATATGAGGCCCCCAGCTGAGAGATATGAGGCCCCCAGCTGAGAGAT 409
DB 6783 CCCCACCTGAGAGATATGAGGCCCCCAGCTGAGAGATATGAGGCCCCCAGCTGAGAGAT 6842

QY 410 GAGCCTCAGCTGAGAGATATGAGGCCCCCAGCTGAGAGATATGAGGCCCCCAGCTGAGAGAT 469
DB 6843 GAGCCTCAGCTGAGAGATATGAGGCCCCCAGCTGAGAGATATGAGGCCCCCAGCTGAGAGAT 6902

QY 470 ATGAGGCCCACTCTAGAGATATGAGGCCCCCAGCTGAGAGATATGAGGCCCCCAGCTCTG 529
DB 6903 ATGAGGCCCACTCTAGAGATATGAGGCCCCCAGCTCTGAGATATGAGGCCCCCAGCTCTG 6962

QY 530 GATATGAGGCCCCCAGCTCTGAGATATGAGGCCCCCAGCTCTGAGATATGAGGCCCCCAGCTCTG 589
DB 6963 GATATGAGGCCCCCAGCTCTGAGATATGAGGCCCCCAGCTCTGAGATATGAGGCCCCCAGCTCTG 7022

QY 590 CGGATACAGAGGCTCAGCTGAGATATGAGGCCCCCAGCTCTGAGATATGAGGCCCCCAGCTCTG 649
DB 7023 CGGATACAGAGGCTCAGCTGAGATATGAGGCCCCCAGCTCTGAGATATGAGGCCCCCAGCTCTG 7082

QY 650 AGGCTCTGAAAAAGAG 666
DB 7083 AGGCTCTGAAAAAGAG 7099

RESULT 4
US-09-864-761-15029

Sequence 15029, Application US/09864761

Patent No. US20020648763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wenheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aecmika-X-1

CURRENT FILING DATE: 2001-05-23

PRIOR FILING DATE: 2000-02-04

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-08-03

PRIOR FILING DATE: 2000-10-04

PRIOR FILING DATE: 2000-10-04

PRIOR FILING DATE: 2000-09-27

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 15029

LENGTH: 471

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

OTHER INFORMATION: MAP TO 299716.4

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2

US-09-864-761-15029

Query Match 13.0%; Score 130.4; DB 9; Length 471;
Best Local Similarity 99.2%; Pred. No. 2.5e-32;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 290 CAGTTATGCTGATGAGGCCCCCAGCTGAGAGATATGAGCCCACTCCGAGATACGAG 349
DB 340 CAGTTATGCTGATGAGGCCCCCAGCTGAGAGATATGAGCCCACTCCGAGATACGAG 399

QY 350 CCCCACCTGAGAGATATGAGGCCCCCAGCTGAGAGATATGAGGCCCCCAGCTGAGAGAT 409
DB 400 CCCCACCTGAGAGATATGAGGCCCCCAGCTGAGAGATATGAGGCCCCCAGCTGAGAGAT 459

QY 410 GAGCCTCAGCTG 421
DB 460 GAGCCTCAGCTG 471

RESULT 5

US-09-764-891-1356

Sequence 1356, Application US/09764891

Publication No. US20030077808A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC006

CURRENT APPLICATION NUMBER: US/09/764,891

PRIOR FILING DATE: 2001-01-17

Prior application data removed - consult PAM or file wrapper

NUMBER OF SEQ ID NOS: 10231

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1356

LENGTH: 321

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

NAME/KEY: SITE

LOCATION: (316)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (321)

OTHER INFORMATION: n equals a,t,g, or c

US-09-764-891-1356

Query Match 11.0%; Score 110.2; DB 11; Length 321;
Best Local Similarity 96.8%; Pred. No. 1e-25;
Matches 120; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGCATTGATCTGATGAGAACCTGCTGTTGAACAACAGATTTGCTGCAACTC 60
DB 195 ATGCATTGATCTGATGAGAACCTGCTGTTGAACAACAGATTTGCTGCAACTC 254

Qy	486	AGGATATGAGCCCACTCTTGGATATAGAAACCCACTCTGGATATGAGCCCAAC	545
Db	388	TGTTTCAACAGCCTCTCTCTCTGTTTCACTAGCCTCTCTCTCTGTTTCAATGAGCCTCTCC	325
Qy	546	TCTCGAATATGAGGCCCACTGCAAGAAATGAAGCCGCGCTGGGGATACAGAGCTC	605
Db	328	TCTCTGTTACGTAGCCTCTCTCTCTCTTTCACATAGCCTCTCTCTGTTTCAACAGCCTC	265
Qy	606	ACCTGCTGATACAGAGCCAGGCTTCAAGAACTTACAGAGCCAGCCTCTGAAAACGA	665
Db	268	TCTCATGTTACGTAGCCTCTCTCTCTCTCATGAGCTATCTCTGTTACGTTGC	208
Qy	666	GCGTTCTCTTCCCTCTGCGCTCTCTCTTCAAGTTACA	701
Db	208	CTCTCTCTCTGTTACATGAGCCTCTCTCTCTCTGTTA	173

```

RESULT 9
US-09-764-891-7571
: Sequence 7571, Application US/09764891
: Publication N. US20030077808A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC006
: CURRENT APPLICATION NUMBER: US/09/764,891
: PRIORITY FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 10231
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 7571
: LENGTH: 291
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-764-891-7571

```

Query Match	9.3%	Score 99.2;	DB 11;	Length 291;
Best Local Similarity	96.9%	Pred. No. 4.3e-20;		
Matches	95;	Conservative	0;	Mismatches 3; Indels 0; Gaps 0

QY	1	ATGCATTGATTCGATGACGAACTCACTGTTGAACAACGATATTTGCTGCAACTTC	60
DB	61	ATGCATTGATTCGATGACGAACTCACTGTTGAACAACGATATTTGCTGCAACTTC	120
QY	61	ATTAGGGAACTATTCAGGAGCTCCATATGGTGCTG	98
DB	121	ATTAGGGAACTATTCAGGAGCTCCATATGGTGATG	158

RESULT 10
 US-09-664-761-4562/c
 : Sequence 4562, Application US/09664761
 : Patent No. US20020648763A1
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Penn, Sharon G.
 : APPLICANT: Rank, David R.
 : APPLICANT: Hanzel, David K.
 :
 : APPLICANT: Chen, Wenhang
 :
 : TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 : TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 :
 : FILE REFERENCE: Aeomica-X-1
 :
 : CURRENT APPLICATION NUMBER: US/09/664,761
 :
 : CURRENT FILING DATE: 2001-05-23
 :
 : PRIOR APPLICATION NUMBER: US 60/180,312
 :
 : PRIOR FILING DATE: 2000-02-04
 :
 : PRIOR APPLICATION NUMBER: US 60/207,456
 :
 : PRIOR FILING DATE: 2000-05-26
 :
 : PRIOR APPLICATION NUMBER: US 09/632,366
 :
 : PRIOR FILING DATE: 2000-08-03

```

1 PRIOR APPLICATION NUMBER: GB 24263.6
2 PRIOR FILING DATE: 2000-10-04
3 PRIOR APPLICATION NUMBER: US 60/236,359
4 PRIOR FILING DATE: 2000-09-27
5 PRIOR APPLICATION NUMBER: PCT/US01/00666
6 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/00667
8 PRIOR FILING DATE: 2001-01-30
9 PRIOR APPLICATION NUMBER: PCT/US01/00664
10 PRIOR FILING DATE: 2001-01-30
11 PRIOR APPLICATION NUMBER: PCT/US01/00669
12 PRIOR FILING DATE: 2001-01-30
13 PRIOR APPLICATION NUMBER: PCT/US01/00665
14 PRIOR FILING DATE: 2001-01-30
15 PRIOR APPLICATION NUMBER: PCT/US01/00668
16 PRIOR FILING DATE: 2001-01-30
17 PRIOR APPLICATION NUMBER: PCT/US01/00663
18 PRIOR FILING DATE: 2001-01-30
19 PRIOR APPLICATION NUMBER: PCT/US01/00662
20 PRIOR FILING DATE: 2001-01-30
21 PRIOR APPLICATION NUMBER: PCT/US01/00661
22 PRIOR FILING DATE: 2001-01-30
23 PRIOR APPLICATION NUMBER: PCT/US01/00670
24 PRIOR FILING DATE: 2001-01-30
25 PRIOR APPLICATION NUMBER: US 60/234,687
26 PRIOR FILING DATE: 2000-09-21
27 PRIOR APPLICATION NUMBER: US 09/608,408
28 PRIOR FILING DATE: 2000-06-30
29 PRIOR APPLICATION NUMBER: US 09/774,203
30 PRIOR FILING DATE: 2001-01-29
31 NUMBER OF SEQ ID NOS: 49117
32 SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1.1
33 SEQ ID NO 4562

```

```

? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AI009176.4
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5.
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
? OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7.
? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7.
? OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
? OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
? OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
? OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.4
? IS-09-864-761-4562

```

Query Match	8.9%	Score 89.4;	DB 9;	Length 1984;
Best Local Similarity	51.6%	Pred. No. 2.7e-18;		
Matches 204; Conservative	0;	Mismatches 191;	Indels 0;	Gaps 0

Qy	300	CTATGGGAGGCCCACTCGTAGATATATGAGCCCACTCTCCGGATATCGAGGCCCACTGCG	359
Db	872	CTCCGAGAGACTCTCTCCAGTTCCGGAGAGACTCTCCAGCTCCAGAGAGACCTCC	813
Qy	360	AGGATATGAGCCCAACCCGTAGAAATGAAGCCCGCTGTGGATATCAGAGCTCAC	419
Db	812	CAGTTCGGAGAGACTCTCTCCAGCTCCGAGAGACTCTCCAGTTCGGAGAGACTCC	755
Qy	420	TGTGGATATGAGGCCCACTCTTGATATGAGAGCCCACTGACAGATATGAGGCC	479
Db	752	TCCCACTCCGAGAGAGACTCTCTCAAGCTCTGAGAGACTCTCCAGTTTGGAGAGAC	693
Qy	480	ACCTTAGATATGAGGCCCACTCTTGATATGAGAACCCCACTCTCGATATGAGC	539
Db	692	ACCTCCAGATTCGGAGAGACTCTCCAGTTTGGAGAGACTCTCCAGTTCGGAGA	633
Qy	540	CCCACTCTCGATATGAGGCCCACTCTGACAGAAATGAGGCCCGCTCGGAGATACAG	599
Db	632	GACTCTCCCAATTCGGAGAGACTCTCCAGTTCGGAGAGACTCTCCCACTCCGG	573

QY 60 AGCCTACACGCTGAAATCAGAGACCAAGGCTCAGAGAAATCTCAGCAGGCCAGGCTCTGGA 659
Db 572 AGAATCTCTCTCCAGTTCCGAGAAATCTCTCCAGTTCTGAGAGAGATCTCTCTATTC 51.3
QY 660 AAACGAGGCTTCTTCTCCCTCTGCGCTCTCTTTC 694
Db 512 CAGAGAGATCTCTCCAGCTCCGAGAGATCTCTC 478

RESULT 11
US-10-029-386-12333/c

```

Sequence 12333, Application US/10029386
Publication No. US2003019470A1
GENERAL INFORMATION:
Applicant: Penn, Sharon G.
Applicant: Rank, David R.
Applicant: Hanzel, David K.
Title of Invention: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
Title of Invention: EXPRESSION ANALYSIS TWO
File Reference: AROMICA-X-2
Current Application Number: US/10/029,386
Current Filing Date: 2001-12-20
Number of Seq ID NOS: 34288
Software: Annomax Sequence Listing Engine vers. 1.1
Seq ID NO 12333

```

Query Match	8.9%	Score 88.8	DB 13	Length 541
Best Local Similarity	55.5%	Pred. No. 1.9e-18		
Matches 171, Conservative		0, Mismatches 137,	Indels 0,	Gaps 0

```

RESULT 12
US-10-029-386-24981
; Sequence 24981, Application US/10029386A
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.

```

```

? TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
? TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
? FILE REFERENCE: ABOMICA-X-2
? CURRENT APPLICATION NUMBER: US/10/029,386
? CURRENT FILING DATE: 2001-12-20
? NUMBER OF SEQ. ID NOS: 34288
? SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
? SEQ. ID NO 24981
? LENGTH: 511
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO Z97205.1
? OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
? US-10-029-386-24981

```

Query Match	8.8%	Score 87.8;	DB 13;	Length 511;
Best Local Similarity	-52.3%	Pred. NO. 3.9e-18;		
Matches 194;	Conservative	0;	Mismatches 177;	Indels 0;
			Gaps	0

```

RESULT 13
US-09-864-761-21308/c
/ Sequence 21308, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wensheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
/ FILE REFERENCE: Aeomica-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ CURRENT FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27

```

PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 21308
LENGTH: 543
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL009178.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
OTHER INFORMATION: EST HUMAN HIT: AM382320.1, EVALUE 8.80e-02
OTHER INFORMATION: SWISSPROT HIT: Q05904, EVALUE 7.00e-12
US-09-864-761-21308

Query Match 8.7%; Score 87; DB 9; Length 543;
Best Local Similarity 54.1%; Pred.No. 7.4e-18;
Matches 177; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

305 GGGCCCCCCTGAGATATGAGCCCACTCCGAGATACGAGCCCACTGAGAT 364
403 GGGAACTCTCTCCGATTCGGAGAGACTTCCAGCTCCGAGAGACTCTCCAGCT 344
365 ATGAGCCCACTCCGTAGAAATGAAGCCCGCTGTGGATACAGAGCTCACTGTGC 424
343 CCGAGAGACTCTCCGATTCGGAGAGACTCTCCAGCTCCAGAGAGACCTCCCA 284
445 GATATGAGCCCACTCTTGTGATACGAGCCCACTCCGAGATATGAGCCCACTC 484
283 GTCCGAGAGACTCTCCGATTCGGAGAGACTCTCCAGCTCCGAGAGACTCTC 224
485 TAGATATGAGCCCACTCTTGTGATATGAGCCCACTCTCGATATGAGCCCACTC 544
223 CCAAGCTCCGAGAGACTCTCAAGCTCTGAGAGAGACTCTCCAGTCTGAGAGAGAC 164
545 CTCTGGATATGAGCCCACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604
163 CTCCAGATTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 104
605 CACCTGTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 631

DB 103 CTCCTCCAGTTCGGAGAGAGACTCTC 77

RESULT 14
US-09-918-995-33536
Sequence 33536, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 33536
LENGTH: 432
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(432)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-33536

Query Match 7.7%; Score 77.2; DB 11; Length 432;
Best Local Similarity 62.4%; Pred.No. 1.2e-14;
Matches 121; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 ATGCCATTGATCTGATGACCACTCACTGTTAACAACGATATTGCTGCAATTC 60
DB 231 ATGCCATTGATCTGATGACCACTGATGATCAAGAGCCCGATATTGTCAAATAC 290
QY 61 ATTAAAGGAATATTCAGGAGCTCCATATGCTGTGGAGAGAGCAAGCTATTAA 120
DB 291 ATCAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 350
QY 121 TTAGCTTCAGAAATGAGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 351 TTGACTTTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 410
QY 181 GCTGTGCGCAGAG 194
DB 411 CAAGCTCCAGAG 424

RESULT 15
US-10-029-386-11265
Sequence 11265, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
FILE REFERENCE: ABOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
PRIOR FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11265
LENGTH: 541
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO Z97205.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
OTHER INFORMATION: SWISSPROT HIT: P22793, EVALUE 1.00e-04

US-10-029-386-11265

Query Match 6.9%; Score 69.2; DB 13; Length 541;
Best Local Similarity 54.3%; Pred. No. 6.1e-12;
Matches 140; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY	308	CCCCACCTGAGATATGAGAGCCCACTCCGATACGAGAGCCCACTGACAGATATG	367
DB	14	CTCCACCTGAGAGCTCTCTCTCACTGAGAGCTCTCTCTCACTGAGAGCTCTT	73
QY	368	GAGCCCAACCCGTAGAAATGAGAGCCCTGTGATACAGAGCTCACTGTGAT	427
DB	74	CTCCTCACTGAGAGGCTCTCTCTCACTGAGAGCTCTCTCTCACTGAGAGCT	133
QY	428	ATGAGAGCCCACTCTTGTATACGAGAGCCCACTGACAGATATGAGAGCCCACTT	487
DB	134	CTCTCTCTCACTGAGAGCTCTCTCTCACTGAGAGCTCTCTCTCACTGAG	193
QY	488	GATATGAGAGCCCACTCTTGTATATGAGAGCCCACTCTGATATGAGAGCCCACTC	547
DB	194	GCTCCCTCTCTCACTGAGAGCTCTCTCTCACTGAGAGCTCTCTCTCACTG	253
QY	548	TCGATATGAGAGCCCACT	565
DB	254	CAGGCTCTCTGAGAGCC	271

Search completed: December 16, 2003, 17:58:45
Job time : 331.43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 09:21:45 ; Search time 2060.06 Seconds

(without alignments)
11809.735 Million cell updates/sec

Title: US-09-864-291-11

Perfect score: 1001

Sequence: 1 atgcctcattgcctgcagac.....atracctcattgcagtag 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 segs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```

Database :
1:  em_estba:*
2:  em_esthum:*
3:  em_estin:*
4:  em_estnu:*
5:  em_estov:*
6:  em_estpl:*
7:  em_estro:*
8:  em_hnc:*
9:  gb_est1:*
10: gb_est2:*
11: gb_hnc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_esthum:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_huv:*
19: em_gss_pln:*
20: em_gss_vtc:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pbg:*
27: em_gss_vtl:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	699.4	69.9	1079	12	BMS64167 AGENCOURT
2	368.2	36.8	10	10	BG699398 602679081
3	362.8	36.2	10	10	BG722815 602655278
4	333.2	33.3	11	11	AK015863 Mus muscu

Result No.	Score	Query Match	Length	DB ID	Description
5	307.8	30.7	431	9	AM169980
6	271.8	27.2	394	9	AL220013
7	263	26.3	701	10	BG701881
8	252.2	25.2	717	14	BY715414
9	247.8	24.8	379	9	AA812559
10	222.2	22.2	814	13	BU961805
11	202.8	20.3	334	9	AA812560
12	187.2	18.7	332	9	AM104932
13	160.4	16.0	488	28	AQ997933
14	138.8	13.9	426	28	AQ997939
15	129.2	12.9	536	12	BM969697
16	125	12.5	403	9	AA468891
17	118.2	11.8	671	29	AG062842
18	105	10.5	632	28	AZ896232
19	104.2	10.4	444	10	BF000754
20	100.6	10.1	488	10	BF476944
21	100.6	10.0	568	12	BI390687
22	100.6	10.0	636	12	BM490729
23	98	9.8	499	13	BM356763
24	97.4	9.7	560	13	BU03740
25	97.4	9.7	700	13	BU284577
26	97.4	9.7	723	9	AJ452591
27	97.4	9.7	744	13	BU230211
28	97.4	9.7	932	13	BU107273
29	90.4	9.0	706	13	BU433499
30	89.8	8.9	723	9	AV899050
31	88.8	8.9	755	13	BM089605
32	88.4	8.8	772	9	AJ454950
33	88	8.8	611	9	AL846063
34	87.6	8.8	686	12	BP021598
35	87.2	8.7	594	12	BM190972
36	87.2	8.7	629	10	BR828725
37	87.2	8.7	816	13	BU916890
38	87.2	8.7	931	13	BU914288
39	86.8	8.7	562	12	BM495739
40	86.6	8.7	622	10	BR468886
41	86.6	8.7	899	10	BR470466
42	86.4	8.6	613	9	AL849211
43	86.4	8.6	635	9	AL889954
44	86.4	8.6	647	9	AL849651
45	86.4	8.6	655	9	AL898723

ALIGNMENTS

RESULT 1
LOCUS BMS64167
DEFINITION AGENCOURT_6560149 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742129
ACCESSION BMS64167
VERSION BMS64167.1 GI:18811738
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 1079)
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: LHAM2759 row: m column: 10
High quality sequence stop: 726.

FEATURES
source

Location/Qualifiers
1. 1079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5742129"
/tissue_type="medulla"
/lab_host="DH10B"
/clone_lib="NIH_MGC_119"
/note="Organ: brain; Vector: pCMV-SPORT6, Site 1: NotI, Site 2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH MGC Library."
BASE COUNT 267 a 297 c 268 g 247 t
ORIGIN

Query Match 69.9%; Score 699.4; DB 12; Length 1079;
Best Local Similarity 96.1%; Pred. No. 6,7e-189;
Matches 749; Conservative 0; Mismatches 26; Indels 4; Gaps 3;
1 ATGCCATTGATCTGATGACGAACCTCACTGTTGAACAACGATTTGCTGCAACTTC 60
269 ATGCCATTGATCTGATGACGAACCTCACTGTTGAACAACGATTTGCTGCAACTTC 328
61 ATTAGGGGAACATTCAGGAGCTCCATTTGGTGGTGGGAAGCAAGCTACTTTTAA 120
339 ATTAGGGGAACATTCAGGAGCTCCATTTGGTGGTGGGAAGCAAGCTACTTTTAA 388
121 TTAGCTTCAGAAATGAGATGCCATTTGATTTGCCAGTTGATGATGATGATGATGAT 180
389 TTAGCTTCAGAAATGAGATGCCATTTGATTTGCCAGTTGATGATGATGATGATGAT 448
181 GCTGTGCCCCGAGATTTTCACCTTAAATGATGATGATGATGATGATGATGATGAT 240
449 GCTGTGCCCCGAGATTTTCACCTTAAATGATGATGATGATGATGATGATGATGAT 508
241 TATGTAATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
509 TATGTAATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 568
301 TATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
569 TATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 628
361 GGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
629 GGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 688
421 GTGGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 479
689 GTGGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 748
480 ACCCTAGAGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 539
749 ACCCTAGAGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 808
540 CCCACCTCTCGGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 599
809 CCCACCTCTCGGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 868
600 AGCCCTCACTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 657
869 AGCCCTCACTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 928
658 GAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 717
929 GAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 988

QY 718 ATGTAA-CCTTGAAGACTCACAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 775
DB 989 ATGTAACTCTTGAAGACTCACAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1047

RESULT 2
BG699398
LOCUS
DEFINITION
602679081P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4811804 5',
mRNA sequence.
ACCESSION
BG699398
KEYWORDS
SOURCES
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Miklos Palikovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/MLML at:
http://image.llnl.gov
Plate: LLAM10703 row: a column: 21
High quality sequence stop: 660.

FEATURES
source
1. 663
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4811804"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI, Site 2: SalI-XhoI (gagag
); Oligo-dT primed using primer 5'-TTTATTTTATTTTATTTT-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 168 a 158 c 169 g 164 t
ORIGIN

Query Match 36.8%; Score 368.2; DB 10; Length 663;
Best Local Similarity 97.1%; Pred. No. 4e-94;
Matches 395; Conservative 0; Mismatches 10; Indels 2; Gaps 2;
1 ATGCCATTGATCTGATGACGAACCTCACTGTTGAACAACGATTTGCTGCAACTTC 60
268 ATGCCATTGATCTGATGACGAACCTCACTGTTGAACAACGATTTGCTGCAACTTC 317
61 ATTAGGGGAACATTCAGGAGCTCCATTTGGTGGTGGGAAGCAAGCTACTTTTAA 120
318 ATTAGGGGAACATTCAGGAGCTCCATTTGGTGGTGGGAAGCAAGCTACTTTTAA 377
121 TTAGCTTCAGAAATGAGATGCCATTTGATTTGCCAGTTGATGATGATGATGATGAT 180
378 TTAGCTTCAGAAATGAGATGCCATTTGATTTGCCAGTTGATGATGATGATGATGAT 437
181 GCTGTGCCCCGAGATTTTCACCTTAAATGATGATGATGATGATGATGATGATGATGAT 240
438 GCTGTGCCCCGAGATTTTCACCTTAAATGATGATGATGATGATGATGATGATGATGAT 497

QY 241 TATGTAATTAATCTGCGGGAAGGGAATATGTGACTCTCCAGATGCTTGTGATTAATGTC 300
 DB 498 TATGTAATTAATCTGCGGGAAGGGAATATGTGACTCTCCAGATGCTTGTGATTAATGTC 557
 QY 301 TATGCGGCCCCACCTGAGGATATGAGCCCACTCCCGATACGAGGCCCACTGCA 360
 DB 558 TATGAGCCCCCACTGAGGATATGAGCCCACTCCCGATACGAGGCCCACTGCA 615
 QY 361 GATATGAGGCCCAACCCGTAGGAATGAAGCCCGCTGTGGATA 407
 DB 616 GATATGAGGCCCAACCCGTAGGAATGAAGCCCGCTGTGGATA 662

RESULT 3
 BG722815 775 bp mRNA linear EST 08-MAY-2001
 LOCUS 602695278F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4827427 5',
 DEFINITION mRNA sequence.
 ACCESSION BG722815 GI:14002002
 VERSION BG722815.1 GI:14002002
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Bakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE NIH-MGC http://mgs.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished
 JOURNAL
 COMMENT Contact: Robert Strauberg, Ph.D.
 Email: cga@nci.nih.gov
 Tissue Procurement: Miklos Palkovics, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LMAN0743 row: 1 column: 20
 High quality sequence stop: 651.
 Location/Qualifiers
 1. 775
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4827427"
 /lab_host="DH10B"
 /clone_id="NIH_MGC_97"
 /note="Organ: testis; Vector: pBluescriptR (modified
 pBluescript KS+); Site 1: BamI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTGTGTTTTTTTNN-3',
 size selected for average insert size 2.2 kb and
 normalized to ROP 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 191 a 212 c 193 g 179 t

ORIGIN

Query Match 36.2%; Score 362.8; DB 10; Length 775;
 Best local similarity 95.1%; Pred. No. 1.6e-92;
 Matches 428; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

QY 1 ATGCAATTTGATCTGATGACCACTCACTGTGAACAACGATATTGCTGCAACTTC 60
 DB 276 ATGCAATTTGATCTGATGACCACTCACTGTGAACAACGATATTGCTGCAACTTC 335
 QY 61 ATTAAGGAATCTATTCAGGAGCTCCCAATAGCTGCGGAAAGGACAGCTACTTTTAA 120
 DB 336 ATTAAGGAATCTATTCAGGAGCTCCCAATAGCTGCGGAAAGGACAGCTACTTTTAA 395

QY 121 TTAGTCTTGAGAAATGAGATGTCATTTGATTTGCCAGTTGATGATGAAAGTCCTCT 180
 DB 396 TTAGTCTTGAGAAATGAGATGTCATTTGATTTGCCAGTTGATGATGAAAGTCCTCT 455
 QY 181 GCTGTGCCGAGAAATTTCACTTAAGACCTTAA-TGACTGTTGAGCTCTATGGAAT 239
 DB 456 GCTGTGCCGAGAAATTTCACTTAAGACCTTAACTGACTGTTGAGCTCTATGGAAT 515
 QY 240 TTATGTAATTAATCTGCGGGAAGGGAATATGTGACTCTCCAGATGCTTGTGATTAATGTC 299
 DB 516 TTATGTAATTAATCTGCGGGAAGGGAATATGTGACTCTCCAGATGCTTGTGATTAATGTC 575
 QY 300 CTATGCGGCCCCACCTGCGAGG-ATATGAGGCCCACTCCCGATACGAGGCCCACTG 358
 DB 576 CTATGAGGCCCACTGCGAGGCTATGAGGCCCACTCCCGATACGAGGCCCACTG 634
 QY 359 CAGATATGAGGCCCAACCCGTAGGAATGAAGCCCGCTGTGGATACAGAGCCCTC-A 417
 DB 635 CAGATATGAGGCCCAACCCGTAGGAATGAAGCCCGCTGTGGATACAGAGCCCTC 694
 QY 418 CCGTGTGGA-TATGAGGCCCACTCTGTG 446
 DB 695 CCGTGTGGA-TATGAGGCCCACTCTGTG 724

RESULT 4
 AK015863 1414 bp mRNA linear HTC 05-DEC-2002
 LOCUS AK015863
 DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
 library, clone:4930521I23 product:hypothetical protein, full insert
 sequence.
 ACCESSION AK015863 GI:12854363
 VERSION AK015863.1 GI:12854363
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE Carninci, P. and Hayashizaki, Y.
 1 High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 2 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 3 Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishibe, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, D.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 TITLE

JOURNAL MEDLINE PUBMED

REFERENCES

AUTHORS

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
 Fleischmann, W., Gaasterland, T., Glessl, C., King, B., Kochiwa, H.,

ACCESSION	AM169980	GI:6401505
VERSION	AM169980.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL	Unpublished	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgsab@remail.nih.gov This clone is available royalty-free through LINT; contact the IMAG Consortium (info@image.lln.gov) for further information. Seq primer: -40UP from Glibco High quality sequence stop: 426.	
FEATURES	Location/Qualifiers	
source	1..431	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/clone="IMAG:2659201"	
	/lab_host="DH10B"	
	/clone_lib="Soares_NFL_T_GBC_S1"	
	/note="Organ: pooled; Vector: pTT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Sco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NDHL19W, testis NHT, and B-cell NCI CGAP GCBI) were mixed, and 88 circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M.A.G.B. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731359. Subtraction by Bento Soares and M. Fatima Bernaldo."	
BASE COUNT	114 a 107 c 82 g 127 t 1 others	
ORIGIN		
Query Match	30.7%; Score 307.8; DB 9; Length 431;	
Best Local Similarity	99.0%; Pred. NO. 6.1e-77;	
Matches 309; Conservative	0; Mismatches 3; Indels 0; Gaps 0;	
Db	690 TTCTCAGGTCCATCTTAAACCTTCAAAATGTAACCTTGAAGACTCAACGAAGAG 749	
	431 TTCTCAGGTCCATCTTAAACCTTCAAAATGTAACCTTGAAGACTCAACGAAGAG 372	
Qy	750 GTACCTTAAATGGAAGTCAGGATTAAGAGACGACGACCTTAGAGTCAATTAATGATC 809	
Db	371 GTACCTTAAATGGAAGTCAGGATTAAGAGACGACGACCTTAGAGTCAATTAATGATC 312	
Qy	810 TGCATTTGGAATTAAGAAACCAATCTCCAGTTTCTGAAGACGACCTTTCTTAAG 869	
Db	311 TGCATTTGGAATTAAGAAACCAATCTCCAGTTTCTGAAGACGACCTTTCTTAAG 252	
Qy	870 GAAGTCAGTACGTGGCAACAGTATGAGAGAGAGAGAGACTCAACCACTAAG 929	
Db	251 GAAGTCAGTACGTGGCAACAGTATGAGAGAGAGAGAGACTCAACCACTAAG 192	
Qy	930 CAGGATTAAGGTTTCCCTTTGTTCACTTTTCACTGTCGCTGAAATGATGATTAACCTC 989	
Db	191 CAGGATTAAGGTTTCCCTTTGTTCACTTTTCACTGTCGCTGAAATGATGATTAACCTC 132	
Qy	990 ATTGTCAGTAG 1001	
Db	131 ATTGTCAGTAG 120	

DEFINITION	9q78ell.11 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE	18413324 3', mRNA sequence.
ACCESSION	A1220013
VERSION	A1220013.1
KEYWORDS	GI:3802216
SOURCE	EST.
ORGANISM	Homo sapiens (human)
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 394)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1264 Std Error: 0.00 Seq primer: -40UP from Gdbco High quality sequence stop: 368. Location/Qualifiers
FEATURES	<p>1..394</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="IMAGE:1841324"</p> <p>/lab_host="DH10B"</p> <p>/clone_id="Soares_NFL_T_GBC_S1"</p> <p>/note="Organ: pooled; Vector: pUT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBH19W, testis NHT, and B-cell NCI CGAP GCBI) were mixed, and as circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1.M.A.G.B. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."</p>
BASE COUNT	101 a 103 c 73 g 117 t
ORIGIN	
Query Match	27.2%; Score 271.8; DB 9; Length 394;
Best Local Similarity	99.3%; Pred. No. 1.2e-66;
Matches 273; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
QY	727 TTGAAGACTCCAGCAAGAGAGTACCCTTAAATTGAAGTCAGAGATTAAGAGACGACT 786
DB	394 TTGAAGACTCCAGCAAGAGAGTACCCTTAAATTGAAGTCAGAGATTAAGAGACGACT 335
QY	787 CAGCTTAGAGTCATTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 846
DB	334 CAGCTTAGAGTCATTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 275
QY	847 TCTCAAGACGACCTTTCTTAAGAGAGTCAGTACGTGGCGAACAGTATGAGAGAAAG 906
DB	274 TCTCAAGACGACCTTTCTTAAGAGAGTCAGTACGTGGCGAACAGTATGAGAGAAAG 215
QY	907 AGAGAGACTCAACCACTAGAGAGCGAGATTAAGTTCCCTGTTCAGCTTTTCAGTGTG 966
DB	214 AGAGAGACTCAACCACTAGAGAGCGAGATTAAGTTCCCTGTTCAGCTTTTCAGTGTG 155
QY	967 TGCTGAGATGATGATTAAGTCACTTCATTCAGAGTAG 1001
DB	154 TGCTGAGATGATGATTAAGTCACTTCATTCAGAGTAG 120
RESULT 7	
LOCUS	BC701881 701 bp mRNA linear EST 07-MAY-2001
DEFINITION	60268323F7L NIH_MGC_95 Homo sapiens cDNA clone IMAGE:18415849 5',

mRNA sequence.
 accession BG701881 GI:13972665
 version BG701881.1
 keywords EST.
 source Homo sapiens (human)
 organism Homo sapiens
 reference Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 title NIH-MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLW at: http://image.llnl.gov
 journal National Institutes of Health, Mammalian Gene Collection (MGC)
 comment Unpublished
 contact: Robert Strausberg, Ph.D.
 email: g99bbs-remail.llnl.gov
 tissue procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA library preparation: Michael J. Brownstein (NHGRI), Shiroki Toshiyuki and Piero Carninci (RIKEN)
 cDNA library arrayed by: The I.M.A.G.E. Consortium (LNLW)
 DNA sequencing by: Incyte Genomics, Inc.
 clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLW at: http://image.llnl.gov
 plate: L1AM10713 row: j column: 10
 high quality sequence stop: 680.
 location/qualifiers
 1. 701
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4815848"
 /tissue_type="hippocampus"
 /lab_host="DH10B"
 /clone_id="NIH MGC 95"
 /note="Organ: brain; Vector: pBluescriptpr (modified pBluescript KS+); Site 1: BamHI, Site 2: SalI-XhoI (gcgag) ; Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3', size selected for average insert size 2.5 kb and normalized to 10⁷ 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NHGRI, National Institutes of Health). Note: this is a NIH_MGC library."
 BASE COUNT 176 a 164 c 189 g 172 t
 ORIGIN
 Query Match 26.3%; Score 263; DB 10; Length 701;
 Best Local Similarity 89.4%; Pred. No. 5.9e-64;
 Matches 387; Conservative 0; Mismatches 30; Indels 16; Gaps 9;
 QY 1 ATGCCATTGATCTGATGACGAACTC-ACGTGTAACCAACGATTTGCT--GCA 55
 DB 261 ATGCCATTGATCTGATGACGAACTCCTTAAGTGAACAACGATTTGCTGCGA 320
 QY 56 ACTTCATTAAAGGAAGTATTCAGGAGCTCCATATGCTGCTGGGAGGAGCAAGTACT 115
 DB 321 CTTCGATTAAAGGAAGTATTCAGGAGCTCCATATGCTGCTGGGAGGAGCAAGTACT 380
 QY 116 TT---AAATTAGCTTGAGAAATGAGATGTCATTGAATTGGCCAGTTGATGCTGAA 171
 DB 381 TTATATAATGTAAGCTTCAAGAAATGAGATGTCATTGAATTGGCCAGTTGATGCTGAA 440
 QY 172 GCTGCTCTGCTGTTGCCGAGAGATTTCCATTAGAACCTTAA-ATGACTGTTGAGCTC 230
 DB 441 GCGGCTCTGCTGCTGCCGAGAGATTTCCATTAGAACCTTAAATGACTGTTGAGCTC 500
 QY 231 TATGGGAATTATGTAATTAATCTGCGGAGGAGAAATATGCA-CTCCACAGATGCTTGT 288
 DB 501 TATGGGCAATTATGTAATTAATCTGCGGAGGAGAAATATGCACTGCCACAGATGCTTGT 560
 QY 289 TGAAGTATATGCTATAGAGGAGCCCACTGAGAGATATGAGGCCCACTCCCGATACGGA 348
 DB 561 ACGATTAT-TGCTATAGAGAGCCCACTGCGGAGATATGAGGCCCACT-CCCGATACGGA 618

QY 349 GCCCACTCTGAGGATATGAGGCCCAACCCGTAGGAATAGAGCCCGCTGTGGATAC 408
 DB 619 GCCCACTCTGAGGATATGAGGCCCAACCCGTAGGAATAGAGCCCGCTGTGGATAC 676
 QY 409 AGAGCTCTACCTG 421
 DB 677 AGAGCTCTACCTG 689
 RESULT 8
 BY715414
 LOCUS
 DEFINITION
 CNA clone 493052123 5', mRNA sequence.
 BY715414
 BY715414.1 GI:27128531
 EST.
 accession BY715414 717 bp mRNA linear EST 17-DEC-2002
 version BY715414
 keywords Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 source Mus musculus (house mouse)
 organism Mus musculus
 reference Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 title NIH-MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLW at: http://image.llnl.gov
 journal National Institutes of Health, Mammalian Gene Collection (MGC)
 comment Unpublished
 contact: Robert Strausberg, Ph.D.
 email: g99bbs-remail.llnl.gov
 tissue procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA library preparation: Michael J. Brownstein (NHGRI), Shiroki Toshiyuki and Piero Carninci (RIKEN)
 cDNA library arrayed by: The I.M.A.G.E. Consortium (LNLW)
 DNA sequencing by: Incyte Genomics, Inc.
 clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLW at: http://image.llnl.gov
 plate: L1AM10713 row: j column: 10
 high quality sequence stop: 680.
 location/qualifiers
 1. 701
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4815848"
 /tissue_type="hippocampus"
 /lab_host="DH10B"
 /clone_id="NIH MGC 95"
 /note="Organ: brain; Vector: pBluescriptpr (modified pBluescript KS+); Site 1: BamHI, Site 2: SalI-XhoI (gcgag) ; Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3', size selected for average insert size 2.5 kb and normalized to 10⁷ 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NHGRI, National Institutes of Health). Note: this is a NIH_MGC library."
 BASE COUNT 176 a 164 c 189 g 172 t
 ORIGIN
 Query Match 26.3%; Score 263; DB 10; Length 701;
 Best Local Similarity 89.4%; Pred. No. 5.9e-64;
 Matches 387; Conservative 0; Mismatches 30; Indels 16; Gaps 9;
 QY 1 ATGCCATTGATCTGATGACGAACTC-ACGTGTAACCAACGATTTGCT--GCA 55
 DB 261 ATGCCATTGATCTGATGACGAACTCCTTAAGTGAACAACGATTTGCTGCGA 320
 QY 56 ACTTCATTAAAGGAAGTATTCAGGAGCTCCATATGCTGCTGGGAGGAGCAAGTACT 115
 DB 321 CTTCGATTAAAGGAAGTATTCAGGAGCTCCATATGCTGCTGGGAGGAGCAAGTACT 380
 QY 116 TT---AAATTAGCTTGAGAAATGAGATGTCATTGAATTGGCCAGTTGATGCTGAA 171
 DB 381 TTATATAATGTAAGCTTCAAGAAATGAGATGTCATTGAATTGGCCAGTTGATGCTGAA 440
 QY 172 GCTGCTCTGCTGTTGCCGAGAGATTTCCATTAGAACCTTAA-ATGACTGTTGAGCTC 230
 DB 441 GCGGCTCTGCTGCTGCCGAGAGATTTCCATTAGAACCTTAAATGACTGTTGAGCTC 500
 QY 231 TATGGGAATTATGTAATTAATCTGCGGAGGAGAAATATGCA-CTCCACAGATGCTTGT 288
 DB 501 TATGGGCAATTATGTAATTAATCTGCGGAGGAGAAATATGCACTGCCACAGATGCTTGT 560
 QY 289 TGAAGTATATGCTATAGAGGAGCCCACTGAGAGATATGAGGCCCACTCCCGATACGGA 348
 DB 561 ACGATTAT-TGCTATAGAGAGCCCACTGCGGAGATATGAGGCCCACT-CCCGATACGGA 618

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Carninci, P., Fukuda
 S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,
 Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno
 H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,
 Nunazaki, R., Ohno, M., Ohnishi, N., Saito, R., Sakazume, N., Sano, H.,
 Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tsgami, M., Takeda, Y.,
 Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct
 Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multipipillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedias: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedias Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
 Location/Qualifiers

FEATURES

Source

1..717
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="4930521123"
 /sex="male"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="SOLR"
 /clone_lib="RIKEN full-length enriched, adult male testis"
 /note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedias Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'-GAGGAGAGAGATCCAGACGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGGAGAGAGCGGCGCCGATTAATCTCGATTAAATTAATCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites."

BASE COUNT

161 a 193 c 198 g 165 t

Query Match 25.2%; Score 252.2; DB 14; Length 717;

Best Local Similarity 74.5%; Pred. No. 7.5e-61;

Matches 333; Conservative 0; Mismatches 108; Indels 6; Gaps 1;

1 ATGCCATTGATCGATGACGACCTCATGTTGAACACGATATTGCGAAACTTC 60
 277 ATGCCATTGATCGATGACGACCTCATGTTGAACACGATATTGCGAAACTTC 336
 61 ATTAAGGAACTATTGAGGAGCTCATATGATGATGAGGAGCAAGTACTTTTAA 120
 337 ATTAAGGAACTATTGAGGAGCTCATATGATGATGAGGAGCAAGTACTTTTAA 396
 121 TTAGCTTCGAGAAATGAGATGCCATTGAATTTGCCAGTTGATGTAAGCTGCTCT 180
 397 ATAGCTTCGAGAAATGAGATGCCATTGAATTTGCCAGTTGATGTAAGCTGCTCT 456
 181 GCTGTGCGCGAGATTTCCATTGAACCTTAATGACGTTCACTATGAGGAATT 240
 457 GCGGTGCGCGAGATTTCCATTGAACCTTAATGACGTTCACTATGAGGAATT 516
 241 TATGTAATTAATGAGGAGAAATATGCACTCCAGATGCTTGTTCAGTTATTTTC 300
 517 TATGTAATTAATGAGGAGAAATATGCACTCCAGATGCTTGTTCAGTTATTTTC 570
 301 TATGAGGCGCCCACTGAGATATGAGACCCCACTCCGAGATGAGAGCCCACTTGA 360
 571 TATGAGGCGCCCACTGAGATATGAGACCCCACTCCGAGATGAGAGCCCACTTGA 630
 361 GATATGAGCCCAACCGGTAGAAATGAGGCCCGCTGTGGGATACAGAGCTCACT 420

Db 631 GGGTATGAGACCCCGACCTGCGGGGTACGAGCGCCCGCGTGTATGTATACCCATCT 690
 Qy 421 GTGCAATGAGACCCCGACCTTGA 447
 Db 691 CCCGATACGATCTCTGCTCTGGA 717

RESULT 9

AA812559/c

LOCUS

DEFINITION

AA812559

VERSION

AA812559.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLES

JOURNAL

COMMENT

AA812559 379 bp mRNA linear EST 31-DEC-1998
 a31c01.s1 Soares_testis_NHT Homo sapiens cDNA clone 1391904 3',
 mRNA sequence.
 AA812559
 AA812559.1 GI:2882623
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 379)
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: CGAP@remail.nih.gov
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
 , Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNL at:
 www.bio.lnl.gov/bhrp/image/image.html
 Insert Length: 1331 Std Error: 0.00
 Seq primer: -40ml3 fwd. RT from Amerham
 High quality sequence stop: 311.
 Location/Qualifiers

FEATURES

Source

1..379
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="1391904"
 /sex="male"
 /lab_host="DH10B"
 /clone_lib="Soares_testis_NHT"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc. and primed with a Not I - oligo(dT) primer [5'-
 TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cos5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

98 a 97 c 68 g 116 t

Query Match 24.8%; Score 247.8; DB 9; Length 379;

Best Local Similarity 99.2%; Pred. No. 8.8e-60;

Matches 249; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

751 TACCTAAATGAGTGCAGATGAGAGACGACTGAGTATGATGATGATCT 810
 379 TACCTAAATGAGTGCAGATGAGAGACGACTGAGTATGATGATGATCT 320
 811 GCATTGGAATTAAGAAACCGAGTCCCATGTTCTCAAGACAGCCCTTTTAAG 870
 319 GCATTGGAATTAAGAAACCGAGTCCCATGTTCTCAAGACAGCCCTTTTAAG 260
 871 AAGTCACTAGTGGCAACAGTATGAGAGAGAGAGAGACTCAACCACTAGAGC 930

Db	259	AAGTCACGACTGTGGGCCAACAGTANTGAGAAGGAAGAGAGAGCTCAACAACACTAAGC	200
Qy	931	AGGATAAGTTTTCCCTTTGTCACGCTTTTGAGTGTCTGTGAATGTGATATTACTCA	990
Db	199	AGGATAAGTTTTCCCTTTGTCACGCTTTTGAGTGTCTGTGAATGTGATATTACTCA	140
Qy	991	TTGTCAAGTAG 1001 	
Db	139	TTGTCAAGTAG 129	
RESULT 10	BUS61805	814 bp mRNA linear EST 21-OCT-2002	
LOCUS	BUS61805	AGENCOURT 10617160 NIH MGC 169 Mus musculus cDNA clone	
DEFINITION		IMAGE:6742447 5', mRNA sequence.	
ACCESSION	BUS61805		
VERSION	BUS61805.1	GI:24191377	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 814) NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished		
AUTHORS	Contact: Robert Strausberg, Ph.D. Email: gsgabp-remail.nih.gov		
TITLE	Liasue Procurement: Dr. Jonathan Kuo, NIMH		
JOURNAL	cDNA Library Preparation: Michael Brownstein Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: LNCM3080 row: 1 column: 06		
COMMENT	High quality sequence steps: 504. Location/Qualifiers 1..814 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:6742447" /lab_host="DH10B (Tl-phage-resistant)" /clone_1lb="NIH_MGC_169" /note="Organ: Testicles; Vector: pDNR-LIB; Site 1: Sfil (ggccatcgagcc); Site 2: Sfil (ggccgcccggcc); cDNA made by oligo-dt priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows: 5'-AAGCATGTGCTATCACGACGAGCGCCATTAAGCGCCGG-3' and 5'-ATTCTAGAGCGCGAGCGGCCGACATG-dT(30)NN-3'. Pull-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC library."		
FEATURES			
source			
BASE COUNT	207 a 238 c 177 g 190 t	2 others	
ORIGIN			
Query Match	22.2%; Score 222.2; DB 13; Length 814;		
Best Local Similarity	67.2%; Pred. No. 3, 2e-52;		
Matches 314;	Conservative 0; Mismatches 153; Indels 0; Gaps 0		
Db	325	GGAGCCCCCACTCCCGAGATACGAGGCCCACTGCAGATATGAGAGCCCAACCCGTAGA	384
Qy	3	GCGTCCTCCCTCTCTCTATATGTATGATACCCCAATGGGCTATGAGATTCCACTCTCGA	62
Db	385	AATGAAGGCCCGCTGTGGATACAGAGCTCACTGTGAGATATGAGGCCCACTCTT	444
Qy	63	TATGAACTTCACCTGTGAGATACGAGATCCCACTCTCGAGATATGAAGCCCCCACTAG	122
Db	445	GGATACGAGAGCCCACCTGACGATATGAGAGGCCCACTGTAGATATGAGGCCCACT	504

DB	123	GAATATGAGACCCCACTCCCAAGTACGAACTAATGGAATCCGGTCTCCACCT	182
QY	505	CTTGAGATGAGAAACCCCACTCTCGATATGAGAGCCCACTCTCGATATGAGAGCCCA	564
DB	183	CCAGATATGAGACCCCGCTTAATGAGGATGAACTCTCTCAAGGCGTGAATCATA	242
QY	565	CTTCGAGGAAATGAAGCCCGCTCGGAGGATACAGAGCTCACTGCTGATCAGAGCC	624
DB	243	CTCTCGAGTCAAGGCGCACTGTGAGCAAGAAAGCTTCACCTCTGGGTCTGAAGCA	302
QY	625	AGGAGCTCAGGATCTACAGAGAGCCAGAGCTCTGAAACAGAGCTCTCTCCCTGCG	684
DB	303	GGCATATCCCATATCTGTGAGCAAGTCCAGATCTCGAATTCAGGCACTTTCCCTGAC	362
QY	685	TCCTCTTCTCAGGTCCATCTTAACCTCTTAAGATGTAACCTTAAGACTCAACAGCA	744
DB	363	TCGCTCTCAAGATCCCACTCTCCCGCTTAAGATGTAACCTTAAGATTCACAGCA	422
QY	745	AAGAGCTAACCTTAATGAAAGTCAGGATGAAGAGCAGCCTCAGCT	791
DB	423	AAACAGCACCTTAATGTTGAAGTCAGATTAACCCGGAACCTTAAGT	469

RESULT	11	334 bp	mRNA	linear	EST	31-DEC-1998
AA812560						
LOCUS	a31c02.g1 Soares_testis_NHT					
DEFINITION	mRNA sequence.					
ACCESSION	AA812560					
VERSION	AA812560.1					
KEYWORDS	EST.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
AUTHORS	1 (base 1 to 334)					
URL	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.					
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index					
JOURNAL	Unpublished					
CONTACT	Contact: Robert Strausberg, Ph.D.					
EMAIL	Email: cgapdb-remail.nih.gov					
LIBRARY	CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.					
CDNA	CDNA Library Arrayed by: Greg Lemmon, Ph.D.					
SEQUENCING	DNA Sequencing by: Washington University Genome Sequencing Center					
CLONE	Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNU at: www-bio.lnlnl.gov/bdtp/image/image.html					
INSERT	Insert Length: 1310 Std Error: 0.00					
SEG	Seg primer: -40ml3 fwd. RT from Amerham					
QUALITY	High quality sequence stop: 189.					
FEATURES	Location/Qualifiers					
SOURCE	1..334					
	/organism="Homo sapiens"					
	/mol_type="mRNA"					
	/db_xref="taxon:9606"					
	/clone="1391906"					
	/sex="male"					
	/lab_host="DH10B"					
	/clone_11b="Soares testis_NHT"					
	/note="Vector: pUT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCCGCCCAATTTTATTTT 3'].					
	Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pUT3D vector. Library went through one round of normalization to Cot5 and was constructed by Bento Soares and M. Fatima Bonaldo."					

BASE COUNT	89 a	86 c	60 g	99 t
ORIGIN				

	Query Match	Score	DB %	Length	334;
	Best local	Similarity	99.0%	Pred. No. 6.4e-47;	
	Matches	204;	Conservative	0;	Mismatches 2; Indels 0; Gaps 0;
QY	796	GTCAATTGATGATCTGATTTGAAAATTAAGAAACCGATGCTCCAGTGTCTCAAGGA	855		
Db	334	GTCAATTGATGATCTGATTTGAAAATTAAGAAACCGATGCTCCAGTGTCTCAAGGA	275		
QY	856	CGACCTTCTTAAAGAAAGTCAAGTACGTGGGCAACAGTATAGAGAAAGAGAGAGAC	915		
Db	274	CAGCTTCTTAAAGAAAGTCAAGTACGTGGGCAACAGTATAGAGAAAGAGAGAGAC	215		
QY	916	TCAACCACTAGACAGAGGATTAAGTTTCCCTGTTCAGCTTTCAAGTCTGCTGAAT	975		
Db	214	TCAACCACTAGACAGAGGATTAAGTTTCCCTGTTCAGCTTTCAAGTCTGCTGAAT	155		
QY	976	GTGATGATTAACCTCAATGTCAAGTNG	1001		
	154	GTGATGATTAACCTCAATGTCAAGTNG	129		

```

RESULT 12
AM104932/c
LOCUS AM104932 322 bp mRNA linear EST 20-OCT-1999
DEFINITION xcd3a02.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2603210 3', mRNA sequence.
ACCESSION AM104932
VERSION AM104932
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1. (bases 1 to 322)
TITLE NCI-CCAP http://www.nccl.nlm.nih.gov/ncicgp.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Glbco
High quality sequence scope: 202.
Location/Qualifiers
1. 322
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2603210"
/lab_host="MDH10B"
/clone_1lb="Soares_NFL_T_GBC_S1"
/note="Organ: pooled; Vector: pTV73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Sco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NDL19W, testis NRT, and B-cell
NCI CGAP GCBI) were mixed, and as circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The divergent
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
1.M.A.G.B. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Patricia Bonaldi."
BASE COUNT 86 a 83 c 55 g 98 t
ORIGIN
Query Match 18.7%; Score 187.2; DB 9; Length 322;
Best Local Similarity 98.4%; Pred. No. 1.8e-42;
Matches 189; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Db 322 TGCATTCTGAAAATTAGAAACGAGTCTCCATGTTCTAGAGACAGCCTTTTAAg 263

Qy 870 GAACTCAGTACCTGCGCAACATGATGAGGAAGAGAGAGACTCAACCACTAGG 929

Db 262 GAATCAGTACCTGCGCAACATGATGAGGAAGAGAGAGACTCAACCACTAGG 203

Qy 930 CAGGATTAAGTTTCCCTGTTCACTTTCACTGTCCTGCTGAATGATGATTAACCTC 989

Db 202 CAGGGAATTAAGTTTCCCTGTTCACTTTCACTGTCCTGGAATGATGATTAACCTC 143

Qy 990 ATTGTCAGTAG 1001

Db 142 ATTGTCAGTAG 131

[illegible]

Matches 243; Conservative 0; Mismatches 112; Indels 6; Gaps 1;

QY 269 TGAATTAATGTTCTATGAGGGGCCCCCACTGAGAGATATAGAGCCCACTCCCGGATACGA 348
DB 457 TCACCGGAGGATACGAGATCCCATCTGAGGATATAGAGCCCACTCCCGGATACGA 398
QY 349 GCGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGGATAC 408
DB 397 GCGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGGATAC 338
QY 409 AGAGCCCTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGGATAC 468
DB 337 GAGAG-----CTGATGATATGAGCTCCCTCCCTCTATATGATAGCTACCCCAATGAGC 284
QY 469 TATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCG 528
DB 283 TATGAGATGATGAGATGAGATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCG 224
QY 529 GATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCG 588
DB 223 GATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCG 164
QY 589 GCGGATATGAGAGCCCTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCC 648
DB 163 ATGGGATCCGGTCCCACTCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCC 104
QY 649 C 649
DB 103 C 103

RESULT 14
A0997939/c 426 bp DNA linear GSS 24-FEB-2000
LOCUS RPCI-23-384A21.TV RPCI-23 Mus musculus genomic clone RPCI-23-384A21
DEFINITION 'genomic survey sequence.'
ACCESSION A0997939
VERSION A0997939.1 GI:7072959
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 426)
Zhao,S., Niernan,W., Feldblum,T., Malek,J., Shatsman,S., Akimret,
B., Levine,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Frazer,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished
Other_GSSes: RPCI-23-384A21.TV
Contact: Shaying Zhao
Department of Bukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buifalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buifalo.edu/orderingframe.htm)
or from Resea.ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/cdb/bac_end/mouse/bac_end_intro.html
Plate: 384 row: A column: 21
Seq primer: SP6
Classes: BAC ends.
Location/Qualifiers
1. 426
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-384A21"

/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1:
Scori, Site 2: Scori; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI methylase. Site
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 88 a 107 c 134 g 97 t
ORIGIN
Query Match 13.9%; Score 138.8; DB 28; Length 426;
Best Local Similarity 66.9%; Pred. No. 1.7e-28;
Matches 230; Conservative 0; Mismatches 107; Indels 7; Gaps 2;
QY 307 GCGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGATAT 366
DB 425 GTCCCATCTGAGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGATAT 366
QY 367 GAGGCCCACTGAGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGATAT 426
DB 365 GTAGCCCATCTGAGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGATAT 312
QY 427 TATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGATAT 486
DB 311 TATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGATAT 252
QY 487 GATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGATAT 546
DB 251 GATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGATAT 192
QY 547 CTCGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCC 605
DB 191 ATGAGATGAGAGCAACTCTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCC 132
QY 606 ACTGCTGATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCC 649
DB 131 ACTGCTGATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCC 88

RESULT 15
BM969697/c 596 bp mRNA linear EST 20-FEB-2003
LOCUS BM969697
DEFINITION UT-CF-DUI-abc-a-10-0-UT.81 UT-CF-DUI Homo sapiens cDNA clone
ACCESSION BM969697
VERSION BM969697.1 GI:19587284
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 596)
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: McCray, FB
McCrays Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research

Genetics (www.reagen.com) or from Open Biosystems

(www.openbiosystems.com).

The following repetitive elements were found in this cDNA

Sequence: 1-29, >AT-rich#low complexity (matched complement) 37-73,

>AT-rich#low complexity (matched complement)

Seq primer: M13 FORWARD

POLYA=Yes.

FEATURES

SOURCE

Location/Qualifiers

1..596

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-CF-DUI-abc-a-10-0-UI"

/tissue_type="Primary Lung Epithelial Cells"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-CF-DUI"

/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a

modified polylinker; Site 1: BcoR I; Site 2: Not I;

UI-CF-DUI is a normalized cDNA library containing the

following tissue(s): Primary Lung Epithelial Cells The

library was constructed according to Bonaldo, Lennon and

Soares, Genome Research, 6:791-806, 1996. First strand

cDNA synthesis was primed with an oligo-dT primer

containing a Not I site. Double stranded cDNA was ligated

to an BcoR I adaptor, digested with Not I, and cloned

directionally into pT73-Pac vector. The oligonucleotide

used to prime the synthesis of first-strand cDNA contains

a library tag sequence that is located between the Not I

site and the (dT)18 tail. The sequence tag for this

library is GGCTGTAGGC.

TAG_Lib=UI-CF-DUI

TAG_Tissue=Lung Epithelial Cells Tissue nos 359-368

TAG_SEQ=GGCTGTAGGC"

BASE COUNT

175 a 86 g 198 t

ORIGIN

137 c

Query Match

Best Local Similarity 12.9%; Score 129.2; DB 12; Length 596;

Matches 134; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY	860	CTTCTTAAGGAGCTAGTGGGCAACAGTATGAGAGAGAGAGAGACTCAA	919
DB	440	CTTACTTCACAACTACAGTACGTGGCAACAGTATGAGAGAGAGAGAGACTCAA	381
QY	920	CCAACTAGAGCAGGATTAAGTTCCCTGTTCACTTTCAAGTGTGCTGAATGGA	979
DB	380	CCAACTAGAGCAGGATTAAGTTCCCTGTTCACTTTCAAGTGTGCTGAATGGA	321
QY	980	TGATTACCTCATTTGTCAAGTAG	1001
DB	320	TGATTACCTCATTTGTCAAGTAG	299

Search completed: December 16, 2003, 17:31:43
Job time : 2064.06 secs

JOURNAL

Patent: WO 0190185-A 11 29-NOV-2001;
 QUEEN'S UNIVERSITY AT KINGSTON (CA); OREGON HEALTH SCIENCES
 UNIVERSITY (US)

FEATURES

source

Location/Qualifiers

1..1001
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 1..708
 /note="unnamed protein product"

CDS

/codon_start=1
 /protein_id="CAD2337.1"
 /db_xref="GI:18675412"
 /translation="MPEDLMTNLTVGPVPAANPIKTIQIAPYGGMEQATPKLVER
 NGDAIRBAQIMVYKASAVANGFPLRTLNDPSSMGIVITBGMCTPQMGCSVIYGG
 APPAGYAPPPGAGPAGYGAQVGNBGPVGRASPVATGAPPLGIGAPYAGP
 PLGYGAPPLGYGAPPLGYGAPPLGYGAPPLGYGAPPLGYGAPPLGYGAPPLGYGAP
 PENBASIPSSASSQVHS"
 BASE COUNT 261 a 254 c 248 g 238 t
 ORIGIN

Query Match

Best Local Similarity 100.0%; Score 705; DB 6; Length 1001;
 Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCATTTGATCTGATGAGAACTCTGTTGAAACACAGATTTTGTGCAAACTTC 60
 DB 1 ATGCCATTTGATCTGATGAGAACTCTGTTGAAACACAGATTTTGTGCAAACTTC 60
 QY 61 ATTAAGGAACTATTGAGGAGCTCCATGATGTTGGTGGGAGGACAAGCTACTTTAAA 120
 DB 61 ATTAAGGAACTATTGAGGAGCTCCATGATGTTGGTGGGAGGACAAGCTACTTTAAA 120
 QY 121 TTAGTCTTGAGAAATGAGATGCCATTGAATTTGCCAGATTGATGTGAAAGCTGCTCT 180
 DB 121 TTAGTCTTGAGAAATGAGATGCCATTGAATTTGCCAGATTGATGTGAAAGCTGCTCT 180
 QY 181 GCTGTGCGCGGAGATTTCCACTTAAAGCTTAAATGACTGTGTTCACTCTTATGGGAATT 240
 DB 181 GCTGTGCGCGGAGATTTCCACTTAAAGCTTAAATGACTGTGTTCACTCTTATGGGAATT 240
 QY 241 TATGTAATTACTGAGGAGGAGGATATGTGCACTCAGAGATGCTTGTGAGTTATTTGC 300
 DB 241 TATGTAATTACTGAGGAGGAGGATATGTGCACTCAGAGATGCTTGTGAGTTATTTGC 300
 QY 301 TATGAGGAGCCCACTGAGATATGAGAGCCCACTCCGAGTACGAGCCCACTGCA 360
 DB 301 TATGAGGAGCCCACTGAGATATGAGAGCCCACTCCGAGTACGAGCCCACTGCA 360
 QY 361 GGATATGAGAGCCCACTGAGATATGAGAGCCCGCTGTGGGATACAGAGCTCACT 420
 DB 361 GGATATGAGAGCCCACTGAGATATGAGAGCCCGCTGTGGGATACAGAGCTCACT 420
 QY 421 GTGGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCA 480
 DB 421 GTGGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCA 480
 QY 481 CCTCTAGATATGAGAGCCCACTCTTGTGATATGAAACCCCACTCTCGATATGAGCC 540
 DB 481 CCTCTAGATATGAGAGCCCACTCTTGTGATATGAAACCCCACTCTCGATATGAGCC 540
 QY 541 CCACCTCTCGATATGAGAGCCCACTGAGATATGAGAGCCCGCTGTGGGATACAGA 600
 DB 541 CCACCTCTCGATATGAGAGCCCACTGAGATATGAGAGCCCGCTGTGGGATACAGA 600
 QY 601 GCGTCACCTGCTGATGAGAGCCAGAGCTCAGAAATCTACAGAGCCCGAGCTGCTGAA 660
 DB 601 GCGTCACCTGCTGATGAGAGCCAGAGCTCAGAAATCTACAGAGCCCGAGCTGCTGAA 660
 QY 661 AACGAGGCTCTCTTCCCTCTGCTCTCTCTCTGAGGCTCATTTCT 705
 DB 661 AACGAGGCTCTCTTCCCTCTGCTCTCTCTCTGAGGCTCATTTCT 705

RESULT 2

BC022546

BC022546 2267 bp mRNA linear PRI 04-FEB-2002
 Homo sapiens, similar to RIKEN cDNA 4930521123 gene, clone

DEFINITION

MGC:26816 IMAGE:481804, mRNA, complete cds.

ACCESSION

BC022546

VERSION

BC022546.1 GI:18490706

KEYWORDS

MGC.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

1 (bases 1 to 2267)
 Strausberg, R.

AUTHORS

Direct Submission
 Submitted (01-FEB-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

TITLE

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
 Toshiyuki and Piero Carninci (RIKEN)

JOURNAL

CDNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) medc@axll.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
 Toshiyuki and Piero Carninci (RIKEN)

COMMENT

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAX Plate: 32 Row: 1 Column: 20
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Similarity but not
 identity to protein.

FEATURES

source

Location/Qualifiers

1..2267
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:26816 IMAGE:481804"
 /tissue_type="Brain, hippocampus"
 /clone_lib="NIH MGC_95"
 /lab_host="DH10B"
 /note="Vector: pBluescript"
 32..961
 /codon_start=1
 /product="Similar to RIKEN cDNA 4930521123 gene"
 /protein_id="AAH22546.1"
 /db_xref="GI:18490707"
 /translation="MAVNSHTRNRGALIPNGESILKSPVVELSPFORSRGVNF
 GRKTGTLFTRSTRVITFSCISDPLSMFPMPLMTNLTVGPVPAANPIKTIQIA
 PYGWSGQATFLVFNNGAIBPAOLMVAASAAAGFPLRTLNDPSSMGIVITG
 GNMCTPQMGCSVIYGGAPPAGYGAQVGNBGPVGRASPVATGAPPLGIGAPYAGP
 PLGYGAPPLGYGAPPLGYGAPPLGYGAPPLGYGAPPLGYGAPPLGYGAPPLGYGAP
 PAGSGAPRHSRTAQA PENBASIPSSASSQVHS"

CDS

BASE COUNT 709 a 477 c 469 g 612 t
 ORIGIN

Query Match

Best Local Similarity 99.1%; Score 698.6; DB 9; Length 2267;
 Matches 701; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGCCATTTGATCTGATGAGAACTCTGTTGAAACACAGATTTTGTGCAAACTTC 60
 DB 254 ATGCCATTTGATCTGATGAGAACTCTGTTGAAACACAGATTTTGTGCAAACTTC 313
 QY 61 ATTAAGGAACTATTGAGGAGCTCCATGATGTTGGTGGGAGGACAAGCTACTTTAAA 120


```

Db      793  CCACCTCTGATGATGAGAGCCAGAGCTCAGAAATCTACAGACGCCAGGCTCTGAA 852
Qy      601  GCCTCAGCTGATGATGAGAGCCAGAGCTCAGAAATCTACAGACGCCAGGCTCTGAA 660
Db      853  GCCTCAGCTGATGATGAGAGCCAGAGCTCAGAAATCTACAGACGCCAGGCTCTGAA 912
Qy      661  AACGAGCTTCTCTCTCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
Db      913  AACGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 957

RESULT 4
HS250D10
LOCUS   220895 bp      DNA      linear      PRI 05-JUN-2003
DEFINITION Human DNA sequence from clone CTA-250D10 on chromosome 22 Contains
            the genes for SREBP2 (sterol regulatory element binding
            transcription factor 2), NAGA (alpha-N-acetylglucosaminidase), a
            gene similar to neuronal-specific septin 3, a pseudogene similar to
            ANT2 (adenine nucleotide translocator 2), 2 mRNAs based on ESTs, a
            genomic marker D2S1178, a CA repeat polymorphism, ESTs and a Cpg
            island, complete sequence.
ACCESSION 299716
VERSION   299716.4  GI:4456457
KEYWORDS  HTG; ANT2; Cpg Island; D2S1178; NAGA; septin 3; SREBP2.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
            Bunkyoja; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 220895)
            Clark, G.
            Direct Submission
            Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
            On Mar 21, 1999 this sequence version replaced gi:1464339.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep CTA-250D10 is
            from the human BAC library described in U-J. Kim et al. (1996)
            Genomics 34, 213-218.
            VECTOR: PBAC108L
            ----- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquerry@sanger.ac.uk
            -----
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            This sequence is the entire insert of clone CTA-250D10 The true
            left end of clone RP1-18601 is at 129979 in this sequence. The true
            left end of clone RP3-359016 is at 1335 in this sequence. The true
            right end of clone RP5-821D11 is at 23458 in this sequence. The
            true right end of clone RP3-359016 is at 118711 in this sequence.

```

```

FEATURES
Source
Location/Qualifiers
1..220895
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="22"
/clone="CTA-250D10"
/clone_lib="CIT978SK-A1"
17..29
/note="2.6 copies 5 mer ATTTG 26% conserved"
33..325
/note="AluX repeat: matches 1..301 of consensus"
326..338
/note="2.6 copies 5 mer TTGCA 26% conserved"
667..683
/note="2.1 copies 8 mer AGAAGACC 34% conserved"
740..751
/note="2.4 copies 5 mer TTTTA 24% conserved"
1443..1457
/note="7.5 copies 2 mer AC 21% conserved"
1715..1997
/note="AluY repeat: matches 5..293 of consensus"
2103..2378
/note="AluY repeat: matches 2..296 of consensus"
2844..2855
/note="12.0 copies 1 mer A 24% conserved"
Complement(2884..3033)
/note="MIR repeat: matches 70..235 of consensus"
Complement(3040..3344)
/note="AluY repeat: matches 1..310 of consensus"
3446..3681
/note="D1M54a repeat: matches 5256..5471 of consensus"
3746..3837
/note="D1M54a repeat: matches 5615..5707 of consensus"
3838..4142
/note="AluY repeat: matches 2..305 of consensus"
4143..4441
/note="D1M54a repeat: matches 5707..5999 of consensus"
4442..4747
/note="AluY repeat: matches 1..292 of consensus"
4748..4858
/note="D1M54a repeat: matches 5999..6110 of consensus"
4881..5172
/note="AluX repeat: matches 1..296 of consensus"
5418..5429
/note="2.4 copies 5 mer GCTGA 24% conserved"
Complement(5769..6021)
/note="match: GSS: Em:A0069839"
6074..6089
/note="2.0 copies 8 mer CCCGGCTG 32% conserved"
6139..6153
/note="2.5 copies 6 mer TTTTA 21% conserved"
6141..6155
/note="3.0 copies 5 mer TTTTA 21% conserved"
Complement(6366..6572)
/note="MIR repeat: matches 5..214 of consensus"
6839..6852
/note="2.3 copies 6 mer GGAAT 28% conserved"
7088..7114
/note="6.8 copies 4 mer TTTA 36% conserved"
7091..7109
/note="1.9 copies 10 mer ATTATTTAC 38% conserved"
Complement(7122..7418)
/note="AluX repeat: matches 1..300 of consensus"
7464..7502
/note="19.5 copies 2 mer AT 78% conserved"
7502..7524
/note="11.5 copies 2 mer AC 46% conserved"
7523..7538
/note="2.0 copies 8 mer CAAATAT 32% conserved"
7535..7563
/note="14.5 copies 2 mer AT 49% conserved"
Complement(7564..7859)

```

	repeat_region	/note="AluJ repeat: matches 1..296 of consensus"
		7873..7893
		/note="5.2 copies 4 mer TGTT 33% conserved"
	repeat_region	7914..8200
		/note="AluJb repeat: matches 1..290 of consensus"
	repeat_region	complement(8202..8318)
		/note="AluJo/FIAM repeat: matches 1..117 of consensus"
	repeat_region	8387..8409
		/note="11.5 copies 2 mer AC 46% conserved"
	repeat_region	8483..8498
		/note="2.0 copies 8 mer GTGGTGGT 32% conserved"
	repeat_region	8486..8496
		/note="2.2 copies 5 mer GTGCT 22% conserved"
	repeat_region	8606..8622
		/note="2.4 copies 7 mer AGAGAAA 34% conserved"
	repeat_region	complement(8767..8983)
		/note="MER115 repeat: matches 33..268 of consensus"
	repeat_region	9004..9014
		/note="2.2 copies 5 mer CTTGG 22% conserved"
	repeat_region	9037..9140
		/note="MIR repeat: matches 17..123 of consensus"
	repeat_region	9509..9823
		/note="AluX repeat: matches 1..312 of consensus"
	repeat_region	9872..9883
		/note="2.4 copies 5 mer GTCCT 24% conserved"
	gene	10071..49606
		/gene="SRBP2"
	mRNA	join(10071..10520,11851..12032,14129..14275,17038..17259
		18577..18691,18783..18964,20469..20661,21182..21363,
		23956..24232,28082..28251,36357..36525,38060..38177,
		40292..40401,41889..42021,43570..43738,46210..46395,
		48103..48214,48680..49606)
		/gene="SRBP2"
		/product="PK25OD10.1 (sterol regulatory element binding
		transcription factor 2)"
		/note="sterol regulatory element binding transcription
		factor 2
		match: cDNA: Em:U02031 Em:U12330 Em:U00968 Em:U09103
		Em:I16995 Em:U2818 Em:U2819 Em:U12329 Em:A0017337
		match: EST: Em:R36312 Em:AL040138 Em:H2808 Em:AA053886"
		/evidence=not experimental
		join((10071..10520,11851..12032,14129..14275,17038..17259,
		18577..18691,18783..18964,20469..20661,21182..21363,
		23956..24232,28082..28251,36357..36525,38060..38177,
		40292..40401,41889..42021,43570..43738,46210..46395,
		48103..48214,48680..48900)
		/gene="SRBP2"
		/note="match: proteins: Sw:Q12772 Sw:O60429 Sw:P36956
		Sw:O60416 Tr:O15352"
		/codon_start=3
		/evidence=not experimental
		/product="PK25OD10.1 (sterol regulatory element binding
		transcription factor 2)"
		/protein_id="CAB41234.1"
		/db_xref="GI:5578715"
		/translation="MLGVSNVGVFPDLPSFOLSPFGSGSSSSSSSSSSSSS
		NGRSSAGVDPSVVSFTVTLPSPFSASAPPAATLVYKSPSTVPPTPATPIT
		PRLPGPQRPQTQLQQQTMTPTSTTGTTRIQPLTYQNATSFYVLQPOYSTLT
		TSVOQPTIIQQOVTVQAQRVLTQTPANSTLQTLAATVAQVAVPVAVQPII
		KTDVLTITTKTDSPVMAAVNPALVTLTPTIALVQVTAAPVVQVGLVPPIM
		GKEVPTIKDVGKVKLEPRKGERHTTHNIIEKYSRKINDKIETLDLVNLTAKM
		HQSVEKRAIDYIKYLQVNHNKLROBNNVNLKIANQSKLSIKDIETLDLVNLTAKM
		DENONVLMSPADSGSDGAPSYISDSRPSPLDDAKVSDSDPDVVALGWMDRS
		RILCVTFICLSFNPLSLTLOWGAHNSDOHPHSGSDSAVSPSSGSGDWMMPT
		LILMVNGVIVLSPVLLVNGEPVIRPHSSSVTFRRHRKQADILDARGPDAALGN
		LQTLAVDGRALPTSRDLACSLSMNVNIRTSLQKRLVRLMKVFQCKRAATPAEAG
		FEDAKTSARDPALAYHRIQLQHLITGKIPASACGDVMNALCAVNAIAEKIKIPST
		LVRIHLTPAMGLKTRCGCKLFILASYFLSPRASLCGPEHSAPVDSLRWLCHPLQCKFF
	CDS	

QY		296	CAGTTATTGTCTATATGGGCCCCCACTGCAGGATAATGAGCCCACTCCCAGATAGAGG	349
Db		170003	CAGTATATGTCTATATGAGCCCCA CTGCAGGATAATGAGCCCACTCCCAGATAGAGG	170062
QY		350	CCCCACTGCAGAGATATGAGGCCAACCGCTGTAAGAAATGAAGCCCGCCTGTGGGATACA	409
Db		170063	CCCCACTGCAGAGATATGAGGCCAACCGCTGTAAGAAATGAAGCCCGCCTGTGGGATACA	170122
QY		410	GAGCTCACCTGTGCGATATYGAGCCCACTCTTGATATACGAGCCCACTGCAGAGAT	469
Db		170123	GAGCTCACCTGTGCGATATYGAGCCCACTCTTGATATACGAGCCCACTGCAGAGAT	170182
QY		470	ATGAGAGCCCCCACTTATGAGATATYGAGCCCACTCTTGATATATGAGACCCTTCCTG	529
Db		170183	ATGAGAGCCCCCACTTATGAGATATYGAGCCCACTCTTGATATATGAGACCCTTCCTG	170242
QY		530	GATATGAGAGCCCCCACTCTCGGATATYGAGCCCACTGCAGAGAAATGAAGCCCGCCTG	589
Db		170243	GATATGAGAGCCCCCACTCTCGGATATYGAGCCCACTGCAGAGAAATGAAGCCCGCCTG	170302
QY		590	CGGATATACAGAGCCTCACTGTGATATACAGAGCCAGCCTCAGAAATCTACAGAGCC	649
Db		170303	CGGATATACAGAGCCTCACTGTGATATACAGAGCCAGCCTCAGAAATCTACAGAGCC	170362
QY		650	AGGCTCTCTGAAAACGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	705
Db		170363	AGGCTCTCTGAAAACGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	170418
RESULT 5				
AX359658				
LOCUS	AX359658	1413 bp	DNA	linear
DEFINITION	Sequence 4 from Patent WO0190185.			PAT 13-FEB-2002
ACCESSION	AX359658			
VERSION	AX359658.1	GI:18675409		
KEYWORDS	unidentified unclassified. unclassified.			
SOURCE	1			
ORGANISM	Oko.R. and Sutoveky, P.			
REFERENCE	Plt2 sperm protein, sperm c-yas, oocyte cytoplasmic c-yes, and uses thereof			
AUTHORS	Patent: WO 0190185-A 4 29-NOV-2001;			
TITLE	QUEBEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES UNIVERSITY (US)			
JOURNAL				
FEATURES	Location/Qualifiers			
source	1..1413			
	/organism="unidentified"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:32644"			
	/note="mammalia"			
	36..977			
	/note="unnamed protein product"			
	/codon_start=1			
	/protein_id="CAD23336.1"			
	/db_xref="gi:18675410"			
	/translation="MAVNSHTSRBRGLIPSGSVLKOCEDVDICPIOKPESTLFLNCTKCKGLTLSSRYVPVPSYSHLVNDPMHLSFPMPPRGLMSCTTBOPFAFNRYIKGTIOAPRGMBQAVFKLSFRKGALIFBAOIAKASAAARGIPLDSVNMTWFTSLGYITIVP GAAVCSSOTPCPAYPIVIYGPBPQYTVQPBYGTTPBPBGYAOPBGYGA PPMGVGAPPVGVPGYGVPGYGA PGYGVPGYGA PGYGVPGYGA PGYGVPGYGA PAG NEALPPAYEAPSAGNTAASHRSMTAQDRTSLPTTSSS"			
CDS	377 a 363 c 369 g 304 c			
BABE COUNT	377 a 363 c 369 g 304 c			
ORIGIN				
Query Match	54.6%; Score 385; DB 6; Length 1413;			
Beat Local Similarity	74.0%; Pred. No. 2.1e-89;			
Matches 538; Conservative	0; Mismatches 165; Indels 24; Gaps 3;			
QY		1	ATGCCATTGATTCGATACGACACTCATCTGTGAAACAACGATATTTGCTGCAGAACTTC	60
Db		258	ATGCCGTTGGCCCTGATATAGTACCTGCACCATTAACAACCAATTTTGGCCCCCACTAC	317

```

Oy 61 ATTAAGGAATATTACAGGACCTCCATATGATGGCTGGGAAGACAAGACTACTTTAA 120
Db 318 ATTAAGGAATATTACAGGACCTCCATATGATGGCTGGGAAGACAAGACTACTTTAA 377
Oy 121 TTAGCTCTCAGAAATGAGATGCCATTGAAATTTGCCAGTTGATGATGAAAGCTCTCT 180
Db 378 TTAGCTCTCAGAAATGAGATGCCATTGAAATTTGCCAGTTGATGATGAAAGCTCTCT 437
Oy 181 GCTGTTCGCCGAGATTTCCACTTGAACCTTAATGATGCTGCTCAGCTTATGAAAT 240
Db 438 GCTGTTCGCCGAGATTTCCACTTGAACCTTAATGATGCTGCTCAGCTTATGAAAT 497
Oy 241 TATGTAATTAATCGGGGAAGGAAT---ATGTCACCTCCACAGATGCTTTGAG----- 292
Db 498 TATGTAATTAATCGGGGAAGGAAT---ATGTCACCTCCACAGATGCTTTGAG----- 557
Oy 293 ---TTATTTGCTATGAGGAGCCCACTGCAAGATATGAGCCCACTTCCGATACGA 348
Db 558 CCAATTTGATCTATGAGCCCACTGCAAGATATGAGCCCACTTCCGATACGA 617
Oy 349 GCGCCACCTGCAAGATATGAGCCCACTGCAAGATATGAGCCCACTTCCGATACGA 408
Db 618 ACTCCACCAAGAAATGAGATGAGCCCACTGCAAGATATGAGCCCACTTCCGATACGA 677
Oy 409 AAGAGCTCACTGTCGATATGAGAGCCCACTGTCATAGAGAGCCCACTTCCGATACGA 468
Db 678 GAGAGCCCTGTCGATATGAGAGCCCACTGTCATAGAGAGCCCACTTCCGATACGA 737
Oy 469 TATGAGCCCACTGTCATAGATATGAGAGCCCACTGTCATAGATATGAGAGCCCACTTCC 528
Db 738 TATGAGCCCACTGTCATAGATATGAGAGCCCACTGTCATAGATATGAGAGCCCACTTCC 797
Oy 529 GATATGAGAGCCCACTGTCATAGATATGAGAGCCCACTGTCATAGATATGAGAGCCCACTT 588
Db 798 GATATGAGAGCCCACTGTCATAGATATGAGAGCCCACTGTCATAGATATGAGAGCCCACTT 857
Oy 589 GGGGAGTACAGAGCTCACTGTCATAGAGAGCCCACTGTCATAGAGAGCTCAAGAGCC 648
Db 858 GCTGGAATGAGAGCTCACTGTCATAGAGAGCTCACTGTCATAGAGAGCTCAAGAGCT 917
Oy 649 -----CAGGCTCTGAAAAGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 696
Db 918 TCTCAGATCTATGACAGCTCAGAGAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCT 977
Oy 697 GTTCATT 703
Db 978 GTTCATT 984

```

```

RESULT 6
BX296515 179222 bp DNA 1linear HTG 26-MAY-2003
LOCUS
DEFINITION
unorderd pieces.
ACCESSION
BX296515.6 GI:31076160
VERSION
HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
Sus scrofa (pig)
SOURCE
Sus scrofa
ORGANISM
Sus scrofa
Mammalia; Eutheria; Cetartiodactyla; Suina, Suidae; Sus.
1 (bases 1 to 179222)
Tracey, A.
Direct Submission
Submitted (25-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On May 26, 2003 this sequence version replaced GI:31043704.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk

```

COMMENT

```

Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: br121d21
----- Summary Statistics
Assembly program: XGAP4, version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 177502 bases at least Q40
Consensus quality: 177971 bases at least Q30
Consensus quality: 178283 bases at least Q20
Insert size: 178722; sum-of-contigs
Insert coverage: 7,82x in Q20 bases; sum-of-contigs quality
quality coverage: 8,59x in Q20 bases; agarose-tp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5850: contig of 5850 bp in length
* 5851 5950: gap of 100 bp
* 74264: contig of 68314 bp in length
* 74265 74364: gap of 100 bp
* 74365 84656: contig of 10292 bp in length
* 84657 84756: gap of 100 bp
* 84757 108493: contig of 23737 bp in length
* 108494 108593: gap of 100 bp
* 108594 114454: contig of 5661 bp in length
* 114455 114554: gap of 100 bp
* 114555 179222: contig of 64668 bp in length.
*
FEATURES
source
1..179222
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9823"
/clone="PigR-121D21"
/clone_11b="PigR"
1..5850
/feature="assembly fragment:02838"
fragment_chain:1"
5951..74264
/feature="assembly fragment:01479"
fragment_chain:1"
74365..84656
/feature="assembly fragment:01113"
fragment_chain:1"
84757..108493
/feature="assembly fragment:00846"
fragment_chain:1"
108594..114454
/feature="assembly fragment:01058"
fragment_chain:2"
114555..179222
/feature="assembly fragment:02416"
fragment_chain:2"
BASE COUNT 44892 a 44880 c 44117 g 44831 t 502 others
ORIGIN
Query Match 30.5% Score 215; DB 2; Length 179222;
Best Local Similarity 73.0%; Pred. No. 6.9e-45;
Matches 294; Conservative 0; Mismatches 100; Indels 9; Gaps 1;
Oy 301 TATGAGGAGCCCACTGCAAGATATGAGAGCCCACTTCCGATACGAGCCCACTTCA 360
Db 55424 TACGAGGCCCACTGCAAGATATGAGAGCCCACTTCCGATACGAGCCCACTTCA 55483
Oy 361 GATATGAGGCCCACTGCAAGATATGAGAGCCCACTTCCGATACGAGCCCACTTCA 420
Db 55484 GATATGAGGCCCACTGCAAGATATGAGAGCCCACTTCCGATACGAGCCCACTTCA 55543

```


Oy	421	GTGGAGATATGAGAGCCCAACCTCTTGATATAGAGAGCCCAACCTGACAGATATAGAGCCCA	480
Db	55544	GAAAGACACAGAGCCCAACGACAGATATAGAGTCCCATCTGCGGGAATATAGAGCCCA	55603
Oy	481	CCTCTAGAAATATGAGAGCCCAACCTCTTGATATAGAACCCCACTCTCGAATATAGAGCC	540
Db	55604	CCTGCAAGATATATAGATGCCCGGCTCCGAGATATAGAGCCCACTGAAAGATATAGAGCC	55653
Oy	541	CCACCTCTCGAATATGAGAGCCCAACCTGACAGAAATAGAGCCCGCTGCGGATACAGA	600
Db	55664	CCCTCTGACGATATATGACACACAAACAGCTGAAATAGAGCCCACTCTAGACATATGAA	55723
Oy	601	GCCCTACCTGCTGATATAGAGAGCCAGGCTTCAGAAATCTACAGACACCCAGGCTCTGAA	660
Db	55724	GCCCCATCTGCTGGAATATAGAGCTGCGCTTCACAAATCTGTGACAGCCAGCC-----	55776
Oy	661	AACGAGCTTCTCTTCCCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	703
Db	55777	--GAGGCTTCTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	55817
RESULT 7			
LOCUS	BX470149	222469 bp	DNA 1linear HNG 07-MAY-2003
DEFINITION	Danio rerio clone DKEX-242K7, *** SEQUENCING IN PROGRESS ***, 50		
ACCESSION	BX470149		
VERSION	BX470149.3	GI:30424228	
KEYWORDS	HNG: HNGS PHASE1.		
SOURCE	Danio rerio (zebrafish)		
ORGANISM	Danio rerio		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Actinopterygii; Neopterygii; Teleostei; Ostariophysi;		
	Cypriniformes; Cyprinidae; Danio.		
	1 (bases 1 to 222469)		
REFERENCE	Burton, J.		
AUTHORS	Direct Submission		
TITLE	Submitted (06-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,		
JOURNAL	Cambridgehire, CB10 1SA, UK. E-mail enquiries:		
	zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk		
	On May 7, 2003 this sequence version replaced gi:30387077.		
COMMENT	----- Genome Center		
	Center: Wellcome Trust Sanger Institute		
	Center code: SC		
	Web site: http://www.sanger.ac.uk		
	Contact: zfish-help@sanger.ac.uk		
	----- Project Information		
	Center project name: ZX242K7		
	----- Summary Statistics		
	Assembly program: XGAP4; version 4.5		
	Chemistry: Dye-terminator; 100% of reads		
	Consensus quality: 197958 bases at least Q40		
	Consensus quality: 206715 bases at least Q30		
	Consensus quality: 212351 bases at least Q20		
	Insert size: 217569; sum-of-contigs		
	Insert size: 165837; 6.2% error; agarose-fp		
	Quality coverage: 2.48x in Q20 bases; sum-of-contigs quality		
	coverage: 3.95x in Q20 bases; agarose-fp		

	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 50 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		

	1 6656: contig of 6656 bp in length		
*	6657 6756: gap of 100 bp		
*	6757 10255: contig of 3499 bp in length		
*	10256 10355: gap of 100 bp		
*	10356 13223: contig of 2868 bp in length		
*	13224 13323: gap of 100 bp		

13324	19935: contig of 6612 bp in length
19336	20035: gap of 100 bp
20036	22597: contig of 2562 bp in length
22598	22697: gap of 100 bp
22698	28569: contig of 5872 bp in length
28670	28669: gap of 100 bp
37966	37965: contig of 9296 bp in length
38066	38065: gap of 100 bp
41141	41140: contig of 3075 bp in length
41241	45108: contig of 3868 bp in length
45109	45208: gap of 100 bp
45209	47730: contig of 2522 bp in length
47731	47830: gap of 100 bp
47831	51884: contig of 4054 bp in length
51885	51984: gap of 100 bp
51985	55557: contig of 3873 bp in length
55558	55957: gap of 100 bp
55958	60498: contig of 4541 bp in length
60499	60598: gap of 100 bp
60599	63641: contig of 3043 bp in length
63642	63741: gap of 100 bp
63742	72674: contig of 8933 bp in length
72675	72774: gap of 100 bp
72775	75602: contig of 2828 bp in length
75603	75702: gap of 100 bp
75703	78062: contig of 2360 bp in length
78063	78162: gap of 100 bp
78163	84485: contig of 6133 bp in length
84486	84585: gap of 100 bp
84586	87553: contig of 2568 bp in length
87554	87653: gap of 100 bp
87654	92487: contig of 4834 bp in length
92488	92587: gap of 100 bp
92588	96104: contig of 3517 bp in length
96105	96204: gap of 100 bp
96205	101479: contig of 5275 bp in length
101480	101579: gap of 100 bp
101580	104979: contig of 3400 bp in length
104980	105079: gap of 100 bp
105080	108003: contig of 2924 bp in length
108004	108103: gap of 100 bp
108104	113151: contig of 5048 bp in length
113152	113251: gap of 100 bp
113252	116585: contig of 3334 bp in length
116586	116685: gap of 100 bp
116686	119346: contig of 2661 bp in length
119347	119446: gap of 100 bp
119447	123125: contig of 3679 bp in length
123126	123225: gap of 100 bp
123226	131266: contig of 8041 bp in length
131267	131366: gap of 100 bp
131367	136193: contig of 4827 bp in length
136194	136293: gap of 100 bp
136294	139482: contig of 3189 bp in length
139483	139582: gap of 100 bp
139583	141875: contig of 2293 bp in length
141876	141975: gap of 100 bp
141976	144122: contig of 2147 bp in length
144123	144222: gap of 100 bp
144223	146968: contig of 2746 bp in length
146969	147068: gap of 100 bp
147069	149399: contig of 2311 bp in length
149400	149499: gap of 100 bp
149500	151963: contig of 4464 bp in length
151964	154053: gap of 100 bp
154054	156579: contig of 2516 bp in length
156580	156679: gap of 100 bp
156680	158982: contig of 2249 bp in length
158929	159028: gap of 100 bp
159029	164542: contig of 5514 bp in length
164543	164642: gap of 100 bp
164643	172054: contig of 7412 bp in length

FEATURES	*	172055	172154:	gap of 100 bp	
	*	172255	175210:	contig of 3056 bp	in length
	*	175211	175310:	gap of 100 bp	
	*	175311	179379:	contig of 4069 bp	in length
	*	179380	179479:	gap of 100 bp	
	*	179480	182698:	contig of 3219 bp	in length
	*	182699	182798:	gap of 100 bp	
	*	182799	186779:	contig of 3981 bp	in length
	*	186800	186879:	gap of 100 bp	
	*	186880	195626:	contig of 8747 bp	in length
	*	195627	195726:	gap of 100 bp	
	*	195727	199396:	contig of 3670 bp	in length
	*	199397	199486:	gap of 100 bp	
	*	206413	206412:	contig of 6916 bp	in length
	*	206513	209619:	contig of 3107 bp	in length
	*	209620	209719:	gap of 100 bp	
*	209720	219083:	contig of 9364 bp	in length	
*	219084	219183:	gap of 100 bp		
*	219184	222469:	contig of 3286 bp	in length	
source	1.	222469	Location/Qualifiers		
	/organism="Danto rexio"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:7955"				
	/clone="DKX1-242K7"				
	/clone_11b="Dantlokey"				
	1.	.6656			
	/note="assembly fragment: 01359				
	fragment chain:1"				
	6757..10255				
	/note="assembly fragment: 01385				
	fragment chain:1"				
	10356..13223				
	/note="assembly fragment: 00592				
	fragment chain:1"				
	13324..19355				
/note="assembly fragment: 00237					
fragment chain:1"					
20036..22597					
/note="assembly fragment: 00949					
fragment chain:1"					
22698..28569					
/note="assembly fragment: 00615					
fragment chain:1"					
28670..37965					
/note="assembly fragment: 01306					
fragment chain:2"					
38066..41140					
/note="assembly fragment: 00487					
fragment chain:2"					
41241..45108					
/note="assembly fragment: 00232					
fragment chain:2"					
45209..47730					
/note="assembly fragment: 00935					
fragment chain:2"					
47831..51864					
/note="assembly fragment: 01415					
fragment chain:2"					
51985..55887					
/note="assembly fragment: 00878					
fragment chain:3"					
55958..60498					
/note="assembly fragment: 01208					
fragment chain:3"					
60599..63641					
/note="assembly fragment: 01266					
fragment chain:3"					
63742..72674					
/note="assembly fragment: 01329					
fragment chain:3"					
72775..75602					

```

/note="assembly_fragment:01183
fragment_chain:4"
m18c_feature 75703..78062
/note="assembly_fragment:01179

Query Match      30.5%; Score 215; DB 2; Length 222469;
Best Local Similarity 73.0%; Pred. No. 7,1e-45;
Matches 294; Conservative 0; Mismatches 100; Indels 9; Gaps 1;

QY      301  TATGGGGCCCCCACTGCAGATATATGAGCCCCCACCCTCCGGATATGAGAGCCCCCACTGCA 360
DB      197080  TACGAGGCCCACTGAAAGACTGGAACCCCAACAGAGGATATGAGAGCCCATCTGG 197139
QY      361  GGATATGAGGCCCAACCCGTAGAGAAATGAGGCCCGCTGTGGATATCAGAGCTCACT 420
DB      197140  GGATATGAGGCTTCACTGCAGATATACCCCTGCTCCAGATATGAGGCTCACT 197199
QY      421  GTGCGATATGAGAGCCCACTCTTGATATAGGAGGCCCACTGAGATATGAGAGCCCA 480
DB      197200  GAAAGACAGAAAGCCCAACAGACAGATATGAGATGCCATTTGGGATATGAGAGCCCA 197259
QY      481  CCTTATGATATGAGAGCCCACTCTTGATATGAGAAACCCCACTCTCGATATGAGACC 540
DB      197260  CTTGCAAGATATGATGATGCCCGCTTCAGATATATGAGGCTCACTGAAAGATATGAGCC 197319
QY      541  CCACCTCTCGATATGAGAGCCCACTCTGAGAGAAATGAGGCCCGCTGCGGATATACA 600
DB      197320  CCTCTGACAGATATGAGACACACACACACTGAAATGAGAGCCCACTTATGAGAA 197379
QY      601  GCCTCACTGCTGATATGAGAGGCCCAAGGCTCAAGAACTTACAGAGAGCCCAAGCTCTGAA 660
DB      197380  GCCCCATCTGCTGAGAAATGAGCTGAGCTTCACAAAATCTTGAGAGGCCAGCC----- 197432
QY      661  AACGAGCTTCTCTTCCCTGCTGCTCCTCTTCTGAGTCAATT 703
DB      197433  --GAGGCTTCTTCTTCCCTTACCTCACTTCTGAGGCCAATT 197473

RESULT 8
BX296540/c 129624 bp DNA linear HTG 02-APR-2003
LOCUS      Sus scrofa clone P1ge-231K18, *** SEQUENCING IN PROGRESS ***, 37
DEFINITION unsorted pieces.
ACCESSION  BX296540
VERSION    BX296540.2 GI:29500961
KEYWORDS  HTG; HTGS PHASE1.
SOURCE    Sus scrofa (pig)
ORGANISM  Sus scrofa
          Bkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
          1 (bases 1 to 129624)
          Burton,J.
          Direct Submission
          Submitted (02-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
          On Apr 2, 2003 this sequence version replaced gi:29335441.
          Genome Center
          Center: Wellcome Trust Sanger Institute
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: humquerry@sanger.ac.uk
          ----- Project Information
          Center project name: br231K18
          ----- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Chemistry: Dye-terminator; 100% of reads
          Consensus quality: 113740 bases at least Q40
          Consensus quality: 118187 bases at least Q30
          Consensus quality: 120712 bases at least Q20
          Insert size: 126024; sum-of-contigs
          Insert coverage: 2.35x in Q20 bases; sum-of-contigs Quality
          Quality coverage: 2.35x in Q20 bases; sum-of-contigs Quality

```


Query Match	29.6%	Score 208.4;	DB 2;	Length 129664;
Best Local Similarity	74.3%	Pred. No. 3,5e-43;		
Matches 263;	Conservative	0;	Mismatches 91;	Indels 0;
			Gaps	0;

RESULT 9	LOCUS	AC113593	168425 bp	DNA	linear	HTG 06-JUN-2007
DEFINITION	Mus musculus clone RP23-363I24,	WORKING DRAFT SEQUENCE, 18 ordered				

ACCESSION	AC113593
VERSION	AC113593.3
KEYWORDS	HTG, HTGS_PHASE2, HTGS_DRAFT, HTGS_FULLTROP.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eulaxyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 168425)
TITLE	Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL	Mus musculus, clone KP23-363124
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 168425)
JOURNAL	Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,

TITLE
JOURNAL
Abstract
Genetic variation in the human genome
Galanzi, J., Gaidyna, S., Ginde, S., Gird, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Kartas, A., Kells, C., Lacroque, K.,
Lamazares, R., Landers, T., Lebochzy, J., Levine, R., Lindblad-Toh, K.,
Litvin, G., Maclean, C., Macdonald, P., Major, J., Margulis, L.,
Matthews, C., McCarthy, K., McGowan, P., McKernan, K., Meldrum, J.,
Meneus, L., Mitsuoka, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phumthang, P., Pierre, N.,
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Roestel, M., Roy, A., Santos, R., Schauer, S.,
Schwaback, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N.,
Stojanovic, N., Strauss, N., Sudirman, A., Talamas, J., Testaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submissions
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gi:121318659.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIMR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L23816
Center clone name: 363_I_24
-----
Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 162227 bases at least Q40
Consensus quality: 164865 bases at least Q30
Consensus quality: 165972 bases at least Q20
Insert size: 162000; agarose-fp
Insert size: 166725; sum-of-contents
Quality coverage: 8.6 in Q20.bases; agarose-fp
Quality coverage: 8.3 in Q20 bases; sum-of-contents
-----

```

```

* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
* 1413 1512: gap of 100 bp in length
* 1513 3366: contig of 1854 bp in length
* 3367 3466: gap of 100 bp in length
* 3467 5542: contig of 2076 bp in length
* 5543 5642: gap of 100 bp in length
* 5643 7949: contig of 2307 bp in length
* 7950 8049: gap of 100 bp in length
* 8050 42510: contig of 34561 bp in length
* 42611 42710: gap of 100 bp in length
* 42711 45717: contig of 3007 bp in length
* 45718 45817: gap of 100 bp in length
* 45818 50106: contig of 4289 bp in length
* 50206 50206: gap of 100 bp in length
* 50207 52945: contig of 2739 bp in length
* 52946 53045: gap of 100 bp in length
* 53046 56269: contig of 3224 bp in length
* 56270 56369: gap of 100 bp in length
* 56370 59897: contig of 3528 bp in length
* 59898 59997: gap of 100 bp in length
* 59998 67207: contig of 7210 bp in length
* 67208 67307: gap of 100 bp in length
* 67308 78743: contig of 11436 bp in length
* 78744 78843: gap of 100 bp in length
* 78844 89991: contig of 11148 bp in length
* 89992 90091: gap of 100 bp in length
* 90092 107918: contig of 17827 bp in length
* 107919 108018: gap of 100 bp in length
* 108019 124874: contig of 16856 bp in length
* 124875 124974: gap of 100 bp in length
* 124975 142093: contig of 17119 bp in length
* 142094 142193: gap of 100 bp in length
* 142194 167211: contig of 24918 bp in length
* 167212 167211: gap of 100 bp in length
* 167212 168425: contig of 1214 bp in length.
Location/Qualifiers
1. 168425
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-363124"
/clone.lib="RPC1-23 Female Mouse BAC"
1. 1412
misc_feature
/note="assembly_fragment"
1513..3366
misc_feature
/note="assembly_fragment"
3467..5542
misc_feature
/note="assembly_fragment"
5643..7949
misc_feature
/note="assembly_fragment"
8050..42610
misc_feature
/note="assembly_fragment"
42711..45717
misc_feature
/note="assembly_fragment"
45818..50106
misc_feature
/note="assembly_fragment"
50207..52945
misc_feature
/note="assembly_fragment"
53046..56269
misc_feature
/note="assembly_fragment"
56370..59897
misc_feature
/note="assembly_fragment"
59998..67207
misc_feature
/note="assembly_fragment"
67308..78743
misc_feature
/note="assembly_fragment"
78844..89991

```

```

/note="assembly_fragment"
90092..107918
misc_feature
/note="assembly_fragment"
108019..124874
misc_feature
/note="assembly_fragment"
124975..142093
misc_feature
/note="assembly_fragment"
142194..167211
misc_feature
/note="assembly_fragment"
167212..168425
/note="assembly_fragment"
vector side:right"
BAC COUNT 44391 a 39572 c 38747 g 44015 t 1700 others
ORIGIN
Query Match 25.4%; Score 179; DB 2; Length 168425;
Best Local Similarity 65.3%; Pred. No. 1.6e-35;
Matches 263; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 303 TGGGGCCCCCACTGAGATATGAGAGCCCACTCCCGATGAGAGCCCACTGCAAG 362
DB 66293 TGGATACGAGACTGTGATATGAGCTCCCTCCCTCTATATGATACCTACCAATGCG 66352
QY 363 ATATGAGCCCAACCCGAGAAATGAAAGCCGCTGTGATATGAGAGCCCACTGCAAG 422
DB 66353 CTATGAGATTCACCTCTCTGATATGAGCTCCACCTGTATGATATGAGATCCCACTCC 66412
QY 423 GCGATATGAGAGCCCACTCTGATATGAGAGCCCACTGCAAGATATGAGAGCCCACT 482
DB 66413 TGGATATGAGAGCCCACTATGATATGAGAGCCCACTCCCAATATGAGATATGAGAG 66472
QY 483 TCTAGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCCCACT 542
DB 66473 TATGAGATCCCGTTCACCACTCCCAATATGAGAGCCCACTGATGAGATGAGATCC 66532
QY 543 ACCCTGAGATATGAGAGCCCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 602
DB 66533 TCCCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 66592
QY 603 CTCACCTGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662
DB 66593 TCACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 66552
QY 663 CGAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
DB 66653 CGAGGACATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 66695
RESULT 10
AC104325 224086 bp DNA linear ROD 30-MAY-2003
LOCUS AC104325
DEFINITION Mus musculus clone rp23-204m3 map 15 strain C57BL/6J, complete
sequence.
ACCESSION AC104325
VERSION AC104325.28 GI:31193955
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 224086)
AUTHORS Jiang,H., Song,L. and Roe,B.A.
TITLE Mus musculus BAC Clone rp23-204m3
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 224086)
REFERENCE Jiang,H., Song,L. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2001) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 224086)
AUTHORS Jiang,H., Song,L. and Roe,B.A.

```

TITLE Direct Submission
JOURNAL Submitted (19-MAR-2003) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 224086)
AUTHORS Jiang, H., Song, L. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-2003) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 5 (bases 1 to 224086)
AUTHORS Jiang, H., Song, L. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (30-MAY-2003) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT ----- Genome Center
Center: Department of Chemistry And Biochemistry
The University Of Oklahoma
Center code: UOKNOR

FEATURES
source 1. 224086
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="15"
/clone="TP23-204m3"
/clone_1lb="RPCL - 23 Female (C57BL/6J) Mouse BAC Library"
BASE COUNT 55592 a 55124 c 56095 g 57275 t
ORIGIN

Query Match 25.4%; Score 179; DB 10; Length 224086;
Best Local Similarity 65.3%; Pred. No 1,7e-35;
Matches 263; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 303 TGGGGCCCACTGACGATATGAGCCCACTCCCGGATACGAGCCCACTTGACAG 362
DB 162839 TGGATACGAGCGTGTGATATGAGCGTCCCTCTCTATATGATGATACCCCAATGGG 162898
QY 363 ATATGAGCCCAACCGGTAGAAATGAAAGCCCGCTGTGATACAGAGCTTCACTGT 422
DB 162899 CTATGAGATGCTCCACTCTCGAATATGAGACTTCACTGTGATGATGAGATCCCACTCC 162958
QY 423 GGATATGAGCCCACTCTTGATATGAGAGCCCACTCGAGATATGAGAGCCCACTCC 482
DB 162959 TGGATATGAGCCCACTATGATATGAGAGCCCACTCCAGATATGAGATCACTACAC 163018
QY 483 TCTAGATATGAGCCCACTCTTGATATGAGAGCCCACTCTCGATATGAGAGCCCACT 542
DB 163019 TATGAGATCGGTTCCCACTCCCAATATGAGAGCCCGCTATGAGGATATGAGAGCTCC 163078
QY 543 ACCTTCGATATGAGAGCCCACTCGAGAAATGAAAGCCCGCTGTGATATGAGAGC 602
DB 163079 TCTTTCAGGCGGATGATATGATATGAGAGCCCACTCTGTGATGATGAGAGAGC 163138
QY 603 CTCACCTGATGATGAGAGCCCACTCGATATGAGAGCCCACTCTGATATGAGAGAGC 662
DB 163139 TCCACCTGCTGGGTCTGAGAGAGCCCACTCATGTCTGTGAGCTCAAGATCTGAAATT 163198
QY 663 CGAGGCTTCTCTCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
DB 163199 CCAGGATCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 163241

RESULT 11
AC107527 253149 bp DNA linear HTG 13-MAY-2003
LOCUS Rattus norvegicus clone CH230-92M24, WORKING DRAFT SEQUENCE, 3
DEFINITION unoriented pieces.
ACCESSION AC107527

VERSION
KEYWORDS AC107527.5 GI:30580771
SOURCE HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
ORGANISM Rattus norvegicus (Norway rat)
Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 253149)
AUTHORS Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alldbrooks, S., Amin, A., Anguitano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhammed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Andrade, C., Dederich, D., Delgado, O., Denson, S., Derramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durkin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Ideblitt, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kovis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshew, L., Louised, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, R., Mawliny, S., McLeod, M. P., McNeill, T. Z., Meenen, B., Milosavljevic, A., Miner, G., Minja, B., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munday, M., Murphy, M., Nat, L., Nankervill, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakenem, O., Okmonu, G., Olampunagoon, A., Pal, S., Parks, K., Pasacker, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poldexter, A., Popovic, D., Prims, B., Pu, L., Puro, M., Qutroz, J., Rachlin, R., Reeves, K., Reiser, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smaj, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svetek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Wilson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, S., Wu, J., Yahub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 253149)
AUTHORS Worley, K. C.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 253149)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 13, 2003 this sequence version replaced gi:23664661.
The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GKFL

Center clone name: CH230-92M24

----- Summary Statistics

Assembly program: Atlas 3.0j

Consensus quality: 242817 bases at least Q40

Consensus quality: 245517 bases at least Q30

Consensus quality: 247516 bases at least Q20

Estimated insert size: 257092; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 250781: contig of 250781 bp in length

250782 250881: gap of unknown length

250882 251966: contig of 1085 bp in length

251967 252066: gap of unknown length

252067 253149: contig of 1083 bp in length.

Location/Qualifiers

1. 253149

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-92M24"

2794. 3331

/note="clone boundary"

clone end: T7

site: Rcor1

end sequence: BH296484"

2829. 3625

/note="clone boundary"

clone end: T7

site: Rcor1

end sequence: BH296484"

complement(246782..247500)

/note="clone boundary"

clone end: Sp6

site: Rcor1

end sequence: BH296486"

BASE COUNT 64040 a 61494 c 61630 g 62235 t 3750 others

ORIGIN

Query Match

Best Local Similarity 64.4%; Pred. No. 2,4e-34;

Matches 261; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

301 TATGGGAGCCACCTGAGATATGAGCCCACTCTCCGATACGAGCCCACTGCA 360

Db 117979 TCTGATATGAGAGCTCGAGATATGATCCCTCTCTCTATACGTACCTACCTTANG 117920
 QY 361 GGTATGAGAGCCCAACCGGTAGAAATGAAAGCCCGCTGTGATACAGAGCTCACT 420
 Db 117919 GGTATGAGAGTCCACCTCTGTGATGAGACCCCACTGTGATAGAGATCCCACT 117860
 QY 421 GTGCGATATGAGAGCCCACTCTGTGATGAGAGCCCACTGTGATATGAGAGCCCA 480
 Db 117859 CCGGATGAGAGAGCCCACTATGAGATGAGAGCCCACTCTCTATACGTACCT 117800
 QY 481 CCTTATGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTGTGATATGAGAGC 540
 Db 117799 CTTATGATATGAGAGTCCCACTCTGTGATATGAGAGCCCACTCTGTGATATGAGAGC 117740
 QY 541 CCACCTCTGTGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTGTGATATGAGAGC 600
 Db 117739 CCGGCTGAGAGAGTATGATATGATATGATATGATATGATATGATATGATATGAGAG 117680
 QY 601 GCTTACCTGTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 Db 117679 GCACTACCTGTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 117620
 QY 661 AAGAGAGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
 Db 117619 TTTTACGAGATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 117575

RESULT 12
 AC132969/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS

1 (bases 1 to 270171)
 Muzny,D,Marie, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsdorfs,S, Amin,A, Anguiano,D, Anyalebech,V, Ayegle,A, Ayodeji,M, Baca,B, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benham,F, Biwaldo,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,B, Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denison,S, Deramo,C, Ding,Y, Dinh,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,L, Evans,K, Egan,A, Baccotto,M, Eugene,C, Evans,C, Falls,T, Fan,G, Fernandez,S, Flintley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Frazer,C,M, Gabriel,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,E, Geer,K, Gill,R, Garcia,M, Gierke,M, Guevara,W, Gunaratne,P, Haaland,M, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogue,M, Hollins,B, Howells,S, Hulyk,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolyet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowls,C, Kraft,C,L, Lebow,H, Lewis,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,M, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenshuber,L, Louised,H, Lozano,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindaratne,M, Mahmoud,M, Mallory,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,B, Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,B, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Parks,K, Pastermak,S, Paul,H, Perez,A, Perez,A, Pfannkuch,C,

FEATURES

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Plapper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzos, M., Quintoz, J., Rachlin, B., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P.,
Rivers, C., Rodkey, T., Rojars, A., Rose, M., Rose, R., Ruiz, S., U.
Sandera, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shartsbeyn, A., Sisson, I., Stiller, C.D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,
Steinle, M., Strong, R., Sutton, A., Swalek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umamji, K.,
Valas, R., Vere, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yeon, L., Yeon, V.,
Yu, F., Zhang, D., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Neidhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 270171)
Rat Genome Sequencing Consortium:
Direct Submission
Submitted (05-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 270171)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:22908490.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/atlas/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly ('a' contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled withNs to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: KBOD
Center clone name: CH230-327L20
----- Summary Statistics -----
Assembly program: Phrap; Version 0.990329
Consensus quality: 251525 bases at least Q40
Consensus quality: 254049 bases at least Q30
Consensus quality: 255593 bases at least Q20
Estimated insert size: 259926; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
270171: 270171 bp in length.
Location/Qualifiers

TITLE The sequence of Homo sapiens BAC clone RP11-45F23
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 148418)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 148418)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 148418)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 148418)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 7 (bases 1 to 148418)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 8, 2000 this sequence version replaced gl:7630812.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics -----
Center project name: H_NH0045F23

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frangon, R., Tatem, M., Cataneese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletier de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-46G22. Actual start of this clone is at base position 1 of RP11-45F23; actual end is at base position 148418 of RP11-45F23.
Location/Qualifiers
1..148418

FEATURES
source

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-45F23"
/clone_1lb="RPC1-11"
1..434
repeat_region
/rpt_family="L1"
507..807
repeat_region
/rpt_family="L1"
807..2585
repeat_region
/rpt_family="L1"
2616..3020
repeat_region
/rpt_family="L1"
3062..3250
repeat_region
/rpt_family="L1"
3251..3596
repeat_region
/rpt_family="MER2_type"
3597..3843
repeat_region
/rpt_family="L1"
9001..9137
repeat_region
/rpt_family="Alu"
11353..11406
repeat_region
/rpt_family="ERVJ"
17191..17869
repeat_region
/rpt_family="L1"
19869..20177
repeat_region
/rpt_family="Alu"
21287..22345
repeat_region
/rpt_family="L1"
22727..22937
repeat_region
/rpt_family="MER2_type"
22933..23133
repeat_region
/rpt_family="MER2_type"
23355..23442
repeat_region
/rpt_family="MIR"
23485..23625
repeat_region
/rpt_family="MER2_type"
24107..24872
repeat_region
/rpt_family="L1"
24873..25320
repeat_region
/rpt_family="ERV1"
25372..25522
misc_feature
/note="similar to EST BC087288 (NTD:98477683)"
28473..28614
repeat_region
/rpt_family="ERV1"
28622..28736
repeat_region
/rpt_family="ERV1"
28790..28843
repeat_region
/rpt_family="ERV1"
28837..29212
repeat_region
/rpt_family="ERV1"
29215..29821
repeat_region
/rpt_family="ERV1"
29823..29972
repeat_region
/rpt_family="ERV1"
29974..30286
repeat_region
/rpt_family="ERV1"
30287..30737
repeat_region
/rpt_family="ERV1"
30745..31216
repeat_region
/rpt_family="L1"
31227..31275
repeat_region
/rpt_family="MIR"
31556..31612
repeat_region
/rpt_family="L2"
32104..32229
repeat_region
/rpt_family="Alu"
32252..32356
repeat_region
/rpt_family="L2"
33845..33935
repeat_region
/rpt_family="MER81"

and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: G1ZG
Center clone name: CH230-203117

----- Summary Statistics

Assembly program: Atlas 3.0;
Consensus quality: 221482 bases at least Q40
Consensus quality: 223612 bases at least Q30
Consensus quality: 224834 bases at least Q20
Estimated insert size: 231669; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbankdraft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 232989: contig of 232989 bp in length
* 232990 233089: gap of unknown length
* 233090 234257: contig of 1168 bp in length
* 234258 234357: gap of unknown length
* 234358 235381: contig of 1024 bp in length.

FEATURES

source

1. 235381
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-203117"

misc_feature

1. 1102
/note="wgs end-extension
clone_end:Sp6"

misc_feature

4245. 4594
/note="clone boundary
clone_end:Sp6
site:Scori"

misc_feature

end_sequence: BH364165"
complement(230331..230677)
/note="clone boundary
clone_end:T7
site:Scori"

misc_feature

231434. 232989
/note="wgs end-extension
clone_end:T7"

BASE COUNT 59674 a 53524 c 55897 g 56470 t 9816 others
ORIGIN

Query Match

Best Local Similarity 58.7%; Pval. No. 2.7e-19;
Matches 202; Conservative 0; Mismatch 142; Indels 0; Gaps 0;

302 ATGGGGCCCAACCGATATGAGCGCCACCTCCGGATACGAGCCCAACCTGACG 361

Db 141772 ATGAGACCTCTCTATTACATGAGACCTCTTATTATACATGAGACCTTCTAT 141713
Cy 362 GATATGAGGCCCAACCGTATGAGAAATGAGAGCCCGCTGAGATACAGCCCTACCTG 421
Db 141712 TACATGAGACCTCTCTATTATACATGAGACCTCTTATTATACATGAGACCTTCTC 141653
Cy 422 TGGATATGAGGCCCAACCGTATGAGAAATGAGAGCCCGCTGAGATATGAGAGCCAC 481
Db 141652 TATTACATGAGACCTCTCTATTATACATGAGACCTCTTATTATACATGAGACCTT 141593
Cy 482 CTCATGATATGAGGCCCAACCGTATGAGAAATGAGAGCCCGCTGAGATATGAGAGCC 541
Db 141592 CTCATATACATGAGACCTCTCTATTATACATGAGACCTCTTATTATACATGAGAC 141533
Cy 542 CACCTCGGATATGAGAGCCCAACCGTATGAGAAATGAGAGCCCGCTGAGATATGAGAG 601
Db 141532 CTCTCATATACATGAGACCTCTCTTATTATACATGAGACCTCTCTATTATACATGAGAT 141473
Cy 602 CCTCACCTGCTGATGAGAGAGCAGGCTTCAGGATCTACAGCA 645
Db 141472 CACTCTCTTATTATACATGAGACCTCTCTATTATACATGAGCA 141429

RESULT 15

AC108000

LOCUS 167587 bp DNA linear PRI 21-MAY-2002
DEFINITION Homo sapiens chromosome 15, clone CTD-2116G1, complete sequence.
AC108000
VERSION AC108000.5 GI:21039854
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE

AUTHORS

TITL

JOURNAL

REFERENCE

AUTHORS

TITL

JOURNAL

REFERENCE

AUTHORS

TITL

JOURNAL

REFERENCE

AUTHORS

TITL

JOURNAL

REFERENCE

AUTHORS

TITL

JOURNAL

REFERENCE

AUTHORS

TITL

JOURNAL

REFERENCE

AUTHORS

TITL

JOURNAL

REFERENCE

AUTHORS

TITL

JOURNAL

REFERENCE

AUTHORS

TITL

JOURNAL

REFERENCE

AUTHORS

TITL

Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menais, L., Minova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., Pollara, V., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vasilev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (28-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 167587)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgeater, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Colymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, U.S., Dodge, S., Fato, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menais, L., Minova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vasilev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
 Submitted (21-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 21, 2002 this sequence version replaced gl:20336146.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L24580
 Center clone name: 2116_G1

FEATURES
source

Location/Qualifiers
 1. 167587
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15"
 /clone="CTD-2116G1"
 /clone_lib="CITD1 Human BAC"
 /complement(11..144)
 /rpt_family="L2"
 repeat_region
 /complement(145..356)
 /rpt_family="AluSg/X"
 repeat_region
 /complement(357..1109)
 /rpt_family="L2"
 repeat_region
 3713..3828
 /rpt_family="L1MB"

repeat_region
 /complement(3834..4006)
 /rpt_family="MIR"
 repeat_region
 4574..4603
 /rpt_family="GABAA)n"
 /complement(4976..5167)
 /rpt_family="MIR"
 repeat_region
 5639..5775
 /rpt_family="L1"
 repeat_region
 6745..6766
 /rpt_family="AT_rich"
 /complement(8250..8370)
 /rpt_family="MIR3"
 repeat_region
 9077..9288
 /rpt_family="MIR"
 repeat_region
 9496..10838
 /rpt_family="L1MD1"
 repeat_region
 10839..11162
 /rpt_family="AluSx"
 /rpt_family="L2"
 repeat_region
 11163..12229
 /rpt_family="L1MD1"
 /complement(13118..13414)
 /rpt_family="AluY"
 repeat_region
 13667..14061
 /rpt_family="L2"
 repeat_region
 14366..14395
 /rpt_family="AT_rich"
 repeat_region
 14748..15052
 /rpt_family="AluSx"
 /complement(15603..15768)
 /rpt_family="MIR"
 repeat_region
 15678..15932
 /complement(15988..16069)
 /rpt_family="MER1"
 repeat_region
 16309..16442
 /rpt_family="MER46A"
 /complement(16557..16730)
 /rpt_family="MIR"
 repeat_region
 16746..16830
 /rpt_family="MIR"
 repeat_region
 16845..16939
 /rpt_family="MER117"
 /complement(17267..17436)
 /rpt_family="MIR3"
 repeat_region
 17914..18014
 /rpt_family="MER103"
 /complement(19124..19415)
 /rpt_family="AluJo"
 repeat_region
 19571..19636
 /rpt_family="L2"
 repeat_region
 19637..19926
 /rpt_family="AluJo"
 /complement(19927..19984)
 /rpt_family="L2"
 repeat_region
 20449..21788
 /rpt_family="L2"
 repeat_region
 21985..22144
 /rpt_family="MIR"
 repeat_region
 22728..22838
 /rpt_family="Char11e8"
 23014..23044
 /rpt_family="AT_rich"
 repeat_region
 23140..23356
 /rpt_family="FLAM_C"
 repeat_region
 23265..23350
 /rpt_family="(TA)n"
 repeat_region
 23357..23381
 /rpt_family="AT_rich"
 repeat_region
 24908..25306
 /rpt_family="AluSx"
 /complement(25805..25979)
 /rpt_family="L1MB4A"
 repeat_region
 /complement(26043..26262)

```
/rpt_family="L1ME4A"  
repeat_region complement(26861..26993)  
/rpt_family="PLAM C"  
repeat_region complement(27023..27317)  
/rpt_family="AluSx"  
repeat_region 27952..28300
```

Query Match 16.1%; Score 113.8; DB 9; Length 167587;
Best Local Similarity 55.1%; Pred. No. 1.6e-18;
Matches 223; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

```
QY 297 TGTCTATGGGGCCCACTGTGAGATATGAGCCCACTCCCGATACGAGCCCAAC 356  
Db 119174 TGTTCACACAGCCTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCAACAGCCTCTCC 119233  
QY 357 TGCAGATATGAGACCCCAACCCGTAGAAATGAAGCCCGCTGTGGATACAGAGCTC 416  
Db 119234 TCCGTTCAGTAGCCTCTCTCTGTTCAACAGCCTCTCTCTGTTCAAGTAGCCTC 119293  
QY 417 ACCGTGCGATATGAGACCCCACTGTGATACGAGCCCACTGACAGATATGAGC 476  
Db 119294 TCCCTCTGTTCAACAGCCTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCAACAGC 119353  
QY 477 CCCACTCTAGATATGAGACCCCACTCTTGATATGAACCCCACTCTCGATATG 536  
Db 119354 CTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCAAGT 119413  
QY 537 AGCCCACTCTCGATATGAGACCCCACTGCAAGAAATGAAGCCCGCTGCGGAT 596  
Db 119414 AGCCTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCA 119473  
QY 597 CAGAGCTCACTGCTGATCAAGAGCAGGCTCAGAACTTACAGAGCCCAAGCTCC 656  
Db 119474 CACAGCTCTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCAACAGCCTCTCTCTGTT 119533  
QY 657 TGAAGAGAGCTCTCTCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 701  
Db 119534 TCATGTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 119578
```

Search completed: December 16, 2003, 15:10:22
Job time : 2626.03 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 05:54:30 ; Search time 188.14 Seconds
(without alignments)
10115.338 Million cell updates/sec

Title: US-09-864-291-11_COPY_1_705

Perfect score: 705
1 atgcattcgtatcgtacgac.....ccctctcagcgcattc 705

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	705	100.0	1001	24	AAS20602
2	446.2	63.3	467	23	AA101225
3	446.2	63.3	467	23	ABL96688
4	414.4	58.8	220895	24	ABK84798
5	411.4	58.4	436	22	ABA67850
6	411.4	58.4	436	22	AAK42003
7	411.4	58.4	436	22	AA148070
8	411.4	58.4	436	24	ABS16034

9	385	54.6	1413	24	AAS20601	DNA encoding bovin
10	375.4	53.2	7099	22	AA104882	Human reproductive
11	375.4	53.2	7099	23	ABL97776	Human testicular a
12	173.6	24.6	894	23	AB70582	DNA encoding novel
13	130.4	18.5	471	22	ABA55246	Human foetal liver
14	130.4	18.5	471	22	AAK28957	Human bone marrow
15	130.4	18.5	471	22	AA134906	Probe #3592 used t
16	130.4	18.5	471	22	ABSO3491	Human genome-deriv
17	110.2	15.6	321	24	AA101355	Human reproductive
18	110.2	15.6	321	23	AB196808	Human testicular a
19	104.2	14.8	2772	23	AA572787	DNA encoding novel
20	104.2	14.8	2772	23	ABA90525	DNA encoding novel
21	102.6	14.6	2243	22	ABA08657	Human extensin hom
22	100	14.2	1824	23	AA581488	DNA encoding novel
23	100	14.2	2850	22	AA579655	DNA encoding novel
24	93.2	13.2	291	22	AA104883	Human reproductive
25	93.2	13.2	291	23	ABL97777	Human testicular a
26	89.4	12.7	1984	22	ABA45936	Human breast cell
27	89.4	12.7	1984	22	ABA56461	Human foetal liver
28	89.4	12.7	1984	22	ABA26096	Probe #4562 for ge
29	89.4	12.7	1984	22	AAK04622	Human brain expres
30	89.4	12.7	1984	22	AAK30133	Human bone marrow
31	89.4	12.7	1984	22	AA114731	Probe #4664 for ge
32	89.4	12.7	1984	22	AA136097	Probe #4783 used t
33	89.4	12.7	1984	22	AA104535	Probe #4526 used t
34	89.4	12.7	1984	22	ABS29782	Human liver single
35	89.4	12.7	1984	24	ABS04714	Human genome-deriv
36	87	12.3	543	22	ABA51054	Human breast cell
37	87	12.3	543	22	ABA69042	Human foetal liver
38	87	12.3	543	22	ABA55988	Probe #14454 for g
39	87	12.3	543	22	AAK17357	Human brain expres
40	87	12.3	543	22	AAK43154	Human bone marrow
41	87	12.3	543	22	AA123925	Probe #13858 for g
42	87	12.3	543	22	AA149232	Probe #17918 used t
43	87	12.3	543	22	AA109525	Probe #9516 used t
44	87	12.3	543	23	ABS42787	Human liver single
45	87	12.3	543	24	ABS17230	Human genome-deriv

ALIGNMENTS

RESULT 1	
AAS20602	
ID	AAS20602 standard; cDNA; 1001 BP.
XX	
AC	AAS20602;
XX	
DT	09-APR-2002 (first entry)
XX	
DB	DNA encoding human testicular WW domain binding protein (htbWP).
XX	
KW	Testicular WW domain binding protein; WBPH; perinuclear chca 32;
KW	PT32; contraceptive; fertility; oocyte activation; vaccine;
KW	globoconfermy; spermiogenesis; spermatoczoa; tyrosine kinase; c-Yes;
KW	immunoreceptor; human; gene; ss.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	primer_bind
FT	1..18
FT	/tag= a
FT	/note= "Binds forward primer AAS20605"
FT	1..708
FT	/tag= b
FT	/product= "htbWP"
FT	/note= "Human testicular WW domain binding protein"
FT	complement (984..1001)
FT	/tag= c
FT	/note= "Binds reverse primer AAS20606"
XX	
PN	WO200190185-A2.
XX	

PD 29-NOV-2001.
 XX
 PF 25-MAY-2001; 2001WO-CA00738.
 XX
 PR 25-MAY-2000; 2000CA-2307128.
 PR 25-MAY-2000; 2000US-206979P.
 XX
 PA (TOOH) UNIV QUEBENS KINGSTON.
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 XX
 PI Olo R, Sutovsky P;
 XX
 DR WPI: 2002-097644/13.
 DR P-PSDB; AAU74610.
 XX
 PT Isolated perinuclear theca 32 polypeptide that interacts with activated
 PT tyrosine kinase c-Yes, for enhancing fertility, treating/diagnosing
 PT diminished fertility and abnormal spermiogenesis and for providing
 PT contraception -
 XX
 PS Claim 62; Fig 4B; 103pp; English.
 XX
 CC The invention describes an isolated perinuclear theca 32 (PT32)
 CC polypeptide (I) which interacts with tyrosine kinase c-Yes. (I) is
 CC useful for: enhancing fertility in a mammal; treating globozoosperm, by
 CC expressing (I) in spermatozoa; inhibiting fertilization, by introducing
 CC (I) or its antigenic fragment into a mammal to elicit an immune
 CC response; enhancing the ability of round spermatozoa to activate oocytes;
 CC treating or diagnosing diminished fertility and abnormal spermiogenesis;
 CC in providing contraceptive agents; identifying contraceptive and
 CC fertility-enhancing agents. The polynucleotide is useful for producing
 CC (I) by recombinant techniques, as vaccine, as diagnostic reagents, and
 CC for chromosome identification. An antibody against (I) is useful in
 CC immunological assays, in immunoneutralization methods, to identify cells
 CC expressing (I), and to purify (I) by affinity chromatography. A
 CC transgenic animal is useful as an animal model for studying human
 CC fertility and reproductive biology, and for screening compounds to
 CC identify modulators of oocyte activation. The use of (I) prevents the
 CC entry of components which are detrimental to embryonic development into
 CC the oocyte during oocyte activation with crude sperm extract and avoids
 CC the propagation of viruses such as HIV (human immunodeficiency virus) and
 CC SIV (simian immunodeficiency virus) carried in the sperm. This sequence
 CC encodes the human testicular WW domain binding protein (hWWBP), described
 CC in the method of the invention.
 CC
 XX
 SQ Sequence 1001 BP; 261 A; 254 C; 248 G; 238 T; 0 other;
 Query Match 100.0%; Score 705; DB 24; Length 1001;
 Best Local Similarity 100.0%; Pred. No. 2.2e-200; Indels 0; Gaps 0;
 Matches 705; Conservative 0; Mismatches 0;

Db 301 TATGGGGCCCCACTGACGATATGAGCCCCCACTCCCGATAGAGCCCACTGCA 360
 Qy 361 GGATATGAGGCCCAACCCCTGTAGAAATGAGAGCCCGCTGTGGATATAGAGCCCTACT 420
 Db 361 GGATATGAGGCCCAACCCCTGTAGAAATGAGAGCCCGCTGTGGATATAGAGCCCTACT 420
 Qy 421 GTGCGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTGTGAGATATGAGAGCCCA 480
 Db 421 GTGCGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTGTGAGATATGAGAGCCCA 480
 Qy 481 CCTTAGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTGTGATATGAGAGCC 540
 Db 481 CCTTAGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTGTGATATGAGAGCC 540
 Qy 541 CCACTCTGTGATATGAGAGCCCACTGTGAGAAATGAGAGCCCGCTGTGGATATGAGCA 600
 Db 541 CCACTCTGTGATATGAGAGCCCACTGTGAGAAATGAGAGCCCGCTGTGGATATGAGCA 600
 Qy 601 GCTGACCTGTGATATGAGAGCCCACTGTGAGAAATGAGAGCCCGCTGTGGATATGAGCA 660
 Db 601 GCTGACCTGTGATATGAGAGCCCACTGTGAGAAATGAGAGCCCGCTGTGGATATGAGCA 660
 Qy 661 AACGAGGCTTCTCTCTCCCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
 Db 661 AACGAGGCTTCTCTCTCCCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
 RESULT 2
 AAL01229
 ID AAL01229 standard; cDNA; 467 BP.
 XX
 AC AAL01229;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DE Human reproductive system related antigen cDNA SEQ ID NO: 1230.
 XX
 KW Human, reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200155320-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01339.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.

PR	14-AUG-2000	2000US-0225570
PR	14-AUG-2000	2000US-0225547
PR	14-AUG-2000	2000US-0225517
PR	14-AUG-2000	2000US-0225558
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226579
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226868
PR	23-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227009
PR	30-AUG-2000	2000US-0226924
PR	01-SEP-2000	2000US-0229387
PR	01-SEP-2000	2000US-0229543
PR	01-SEP-2000	2000US-0229335
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	06-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0233080
PR	08-SEP-2000	2000US-0233081
PR	12-SEP-2000	2000US-0233196
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0234401
PR	14-SEP-2000	2000US-0234400
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234224
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0235327
PR	29-SEP-2000	2000US-0235367
PR	29-SEP-2000	2000US-0235368
PR	29-SEP-2000	2000US-0235369
PR	29-SEP-2000	2000US-0235370
PR	02-OCT-2000	2000US-0235802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237039
PR	13-OCT-2000	2000US-0239335
PR	13-OCT-2000	2000US-0239337
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0244765
PR	08-NOV-2000	2000US-0244766
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527

PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246533.
 PR 08-NOV-2000; 2000US-0246593.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-465570/50.
 XX P-PSDB; AAM95259.
 XX
 PT Isolated nucleic acid molecule encoding a reproductive system antigen -
 XX is used in preventing, treating or ameliorating a medical condition -
 XX
 PS Claim 1; SEQ ID NO 1230; 1297bp + Sequence Listing; English.
 XX
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a coding sequence of the
 CC invention.
 XX
 SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;

	Query Match	Similarity	Score	446.2;	DB	22;	Length	467;
	Best Local	Similarity	98.5%;	Pred.	No.3.8e-123;			
	Matches	459;	Conservative	1;	Indels	5;	Gaps	1;
Qy	129	CAGAAATGAGATGCCATTGAAATTTGGCCAGATTGATGATGAAAGCTGCTGCTGTGC	188					
Db	2	CAGAAATGAGATGCCATTGAAATTTGGCCAGATTGATGATGAAAGCTGCTGCTGTGC	61					
Qy	189	CCGAGATTTCCACTTGAACCTTAAATGATGTGTTGAGCTCTATGGAATTTATGTAAT	248					
Db	62	CCGAGATTTCCACTTGAACCTTAAATGATGTGTTGAGCTCTATGGAATTTATGTAAT	121					
Qy	249	TACTGGGAAAGGATATGTGCACCTCCACAGATGCTGTATCAATATGTTATGAGGC	308					

DB 122 TACTGGGAGGAAATATGTGCACTCCACAGATGCTTGTTCAGTTATGTCTATGAGC 181
QY 309 CCACCTGAGAGATATGAGCCCCACCTCCGGATACGAGCCCCACCTGACAGATATG 368
DB 182 CCACCTGAGAGATATGAGCCCCACCTCCGGATACGAGCCCCACCTGACAGATATG 241
QY 369 AGCCCAACCCGTAAGAAATGAAAGCCCCGCTGTGGATACAGAGCTTCACTGTGGATA 428
DB 242 AGCCCAACCCGTAAGAAATGAAAGCCCCGCTGTGGATACAGAGCTTCACTGTGGATA 301
QY 429 TGGAGGCCCACTCTTGATACGAGCCCCACCTGAGAGATATGAGCCCCACCTCTAG 488
DB 302 TGGAGGCCCACTCTTGATACGAGCCCCACCTGAGAGATATGAGCCCCACCTCTAG 361
QY 489 ATATGAGGCCCC-ACCTCTTGATATGAAACCCCACTCTGGATATGAGCCCCACCTC 547
DB 362 ATATGAGGCCCCACCTCTTGATATGAAACCCCACTCTGGATATGAGCCCCACCTC 421
QY 548 TCGATATGAGCCCCCACTGACAGAAATGAAAGCCCCGCTGTGGG 593
DB 422 TCGATATGAGCCCCCACTGACAGAAATGAAAGCCCCGCTGTGGG 467

RESULT 3
ABL96688
ID ABL96688 standard; cDNA; 467 BP.
XX
AC ABL96688;
XX
XX 21-JUN-2002 (first entry)
DT
XX
DE Human testicular antigen encoding cDNA SEQ ID NO: 356.
XX
XX
KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
XX
PN MO200155317-A2.
PD
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214866.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217467.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0232404.
PR 14-SEP-2000; 2000US-0232423.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 12-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.

PR	08-NOV-2000;	2000US-0246609.	
PR	08-NOV-2000;	2000US-0246610.	
PR	08-NOV-2000;	2000US-0246611.	
PR	08-NOV-2000;	2000US-0246612.	
PR	08-NOV-2000;	2000US-0246613.	
PR	17-NOV-2000;	2000US-0249207.	
PR	17-NOV-2000;	2000US-0249208.	
PR	17-NOV-2000;	2000US-0249209.	
PR	17-NOV-2000;	2000US-0249210.	
PR	17-NOV-2000;	2000US-0249211.	
PR	17-NOV-2000;	2000US-0249212.	
PR	17-NOV-2000;	2000US-0249213.	
PR	17-NOV-2000;	2000US-0249214.	
PR	17-NOV-2000;	2000US-0249215.	
PR	17-NOV-2000;	2000US-0249216.	
PR	17-NOV-2000;	2000US-0249217.	
PR	17-NOV-2000;	2000US-0249218.	
PR	17-NOV-2000;	2000US-0249244.	
PR	17-NOV-2000;	2000US-0249245.	
PR	17-NOV-2000;	2000US-0249264.	
PR	17-NOV-2000;	2000US-0249265.	
PR	17-NOV-2000;	2000US-0249297.	
PR	17-NOV-2000;	2000US-0249299.	
PR	17-NOV-2000;	2000US-0249300.	
PR	01-DEC-2000;	2000US-0250160.	
PR	01-DEC-2000;	2000US-0250391.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	06-DEC-2000;	2000US-0256719.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259676.	
PA	(HUMA-)	HUMAN GENOME SCI INC.	
PI	Rosen CA,	Barash SC, Ruben SM;	
PI	WPI;	2001-483332/52.	
DR	Nucleic acids encoding	973 human testicular antigen polypeptides,	
XX	useful for preventing,	diagnosing and/or treating testicular cancer	-
XX	Claim 1;	SEQ ID NO 356; 766bp; English.	
XX	The present invention	provides the protein and coding sequences of 973	
XX	human testicular antigens,	and fragments of their genomic sequences. The	
XX	sequences can be used	in the treatment of cardiovascular, urinary system,	
XX	reproductive system,	immune, respiratory, neurological and	
XX	gastrointestinal disorders,	infections, and particularly cancer,	
XX	especially testicular	cancers. The present sequence is a cDNA of the	
XX	invention.		
XX	Sequence	467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;	
XX	Query Match	63.3%; Score 446.2; DB 23; Length 467;	
XX	Best local Similarity	98.5%; Pred. No. 3.8e-123;	
XX	Matches 459; Conservative	1; Mismatches 5; Indels 1; Gaps	1
QY	129	CAGAAATGAGAGTGCATTGATTTGGCCAGTGTGATGTGGAAGCTGCTCTGCTGTTC	188
DB	2	CAGAAATGAGAGTGCATTGATTTGGCCAGTGTGATGTGGAAGCTGCTCTGCTGTTC	61
QY	189	CCGAGGATTTCCACTTAAAGCTTAAATGACTGTTCACTCTATGAGATTTATGAT	248
DB	62	CCGAGGATTTCCACTTAAAGCTTAAATGACTGTTCACTCTATGAGATTTATGAT	121
QY	249	TATCGGGGAAAGGATTTATGTGACGCCAGATGCTGTTCAGTATTTGTCTATGAGGC	308
DB	122	TATCGGGGAAAGGATTTATGTGACGCCAGATGCTGTTCAGTATTTGTCTATGAGGC	181

QY	309	CCGACCTGACAGATATATGAGCCCACTCCGGATACGAGGCCCACTGGACGATATGG	368
Db	182	CCGACCTGACAGATATATGAGGCCCACTCCGGATACGAGGCCCACTGGACGATATGG	241
QY	369	AGCCCAACCCCTGAGAAATGAAAGCCCGCTGTGGATACAGAGCTCACTGTGCATAT	428
Db	242	AGCCCAACCCCTGAGAAATGAAAGCCCGCTGTGGATACAGAGCTCACTGTGCATAT	301
QY	429	TGAGAGCCCACTCTTGGATACGAGGCCCACTGAGAGATATGAGAGCCCACTCTTAG	488
Db	302	TGAGAGCCCACTCTTGGATACGAGGCCCACTGAGAGATATGAGAGCCCACTCTTAG	361
QY	489	ATATGAGAGCCCACTCTTGGATATGGAACCCCACTCTGGATATGAGAGCCCACTCT	547
Db	362	ATATGAGAGCCCACTCTTGGATATGGAACCCCACTCTCTGGAATATGAGAGCCCACT	421
QY	548	TCGATATATGAGAGCCCACTGACAGAAATGAAAGCCCGCTGTGGAG	593
Db	422	TCGATATATGAGAGCCCACTGACAGAAATGAAAGCCCGCTGTGGAG	467
RESULT 4			
ID	ABK84798	standard, cDNA, 220895 BP.	
XX	AC	ABK84798;	
XX	DT	14-AUG-2002 (first entry)	
DE	XX	Human CDNA differentially expressed in granulocytic cells #1369.	
XX	XX	Human; ss; granulocytic cell; DNA chip; bacterial infection;	
KW	KW	viral infection; paratyphoid infection; protozoal infection;	
KW	KW	fungal infection; sterile inflammatory disease; psoriasis;	
KW	KW	rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;	
KW	KW	cardiac reperfusion injury; renal reperfusion injury; ARDS;	
KW	KW	adult respiratory distress syndrome; inflammatory bowel disease;	
KW	KW	Crohn's disease; ulcerative colitis; periodontal disease;	
KW	KW	granulocyte activation; chronic inflammation; allergy.	
OS	OS	Homo sapiens.	
XX	PN	WO200228999-A2.	
XX	XX	11-APR-2002.	
PD	XX		
XX	PF	03-OCT-2001; 2001WO-US30821.	
XX	PR	03-OCT-2000; 2000US-237189P.	
XX	PA	(GENE-) GENE LOGIC INC.	
XX	PI	Beazer-Barclay Y, Weisman SM, Yamaga S, Vockley J;	
DR	XX	WPI; 2002-435328/46.	
XX	PT	Detecting granulocyte activation by detecting differential expression	
PT	PT	of genes associated with granulocyte activation, which serves as	
PT	PT	diagnostic markers that is useful for monitoring disease states and	
PT	PT	drug toxicity -	
PS	XX	Claim 1; SEQ ID No 1369; 114pp; English.	
XX	XX		
CC	CC	The invention relates to detecting (M1) granulocyte (GC) activation	
CC	CC	(GCA), by detecting the level of expression of gene(s) (Gs) identified by	
CC	CC	DNA chip analysis as given in the specification, and comparing	
CC	CC	the expression level to an expression level in an unactivated	
CC	CC	GC, where differential expression of Gs is indicative of GCA.	
CC	CC	Also included are modulating (M2) GA by contacting GC with an agent	
CC	CC	that alters the expression of at least one gene in Gs; (2) screening (M3)	
CC	CC	for an agent capable of modulating GCA or an inflammation (especially	
CC	CC	chronic) in a tissue, an allergic response in a subject, exposure of a	

CC	RESULT 4
XX	ABR84798
ID	ABR84798 standard; cDNA, 220895 BP.
XX	
AC	ABR84798;
XX	
DT	14-AUG-2002 (first entry)
XX	
DE	Human cDNA differentially expressed in granulocytic cells #1369.
XX	
KM	Human; ss; granulocytic cell; DNA chip; bacterial infection;
KM	viral infection; parasitic infection; protozoal infection;
KM	fungal infection; sterile inflammatory disease; psoriasis;
KM	rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KM	cardiac reperfusion injury; renal reperfusion injury; ARDS;
KM	adult respiratory distress syndrome; inflammatory bowel disease;
KM	Crohn's disease; ulcerative colitis; periodontal disease;
KM	granulocyte activation; chronic inflammation; allergy.
XX	
OS	Homo sapiens.
XX	
FN	WO200228999-A2.
XX	
PD	11-APR-2002.
XX	
PE	03-OCT-2001; 2001WO-US30821.
XX	
PR	03-OCT-2000; 2000US-237189P.
XX	
PA	(GENE-) GENE LOGIC INC.
XX	
PI	Beazer-Barclay Y, Weisman SM, Yamaga S, Vockley J;
DR	WPI; 2002-435328/46.
XX	
PT	Detecting granulocyte activation by detecting differential expression
PT	of genes associated with granulocyte activation, which serves as
PT	diagnostic markers that is useful for monitoring disease states and
PT	drug toxicity -
XX	
PS	Claim 1; SEQ ID NO 1369; 114pp; English.
XX	
CC	The invention relates to detecting (M1) granulocyte (GC) activation
CC	(GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC	DNA chip analysis as given in the specification, and comparing
CC	the expression level to an expression level in an unactivated
CC	GC, where differential expression of Gs is indicative of GCA.
CC	Also included are modulating (M2) GA by contacting GC with an agent
CC	that alters the expression of at least one gene in Gs; (2) screening (M3)
CC	for an agent capable of modulating GCA or an inflammation (especially
CC	chronic) in a tissue, an allergic response in a subject, exposure of a


```

RESULT 6
AAK42003
ID AAK42003 standard; DNA; 436 BP.
AC
XX AAK42003;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 16560.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN M0200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001MO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 16560; 658bp + Sequence Listing; English.
XX
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
Query Match 58.4%; Score 411.4; DB 22; Length 436;
Best Local Similarity 99.8%; Pred. No. 9.3e-113;
Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 293 TTATTGCTATGAGGCCCCCACTGCAAGATATGAGGCCCACTCCCGATACGAGCCC 352
DB 1 TTATTGCTATGAGGCCCCCACTGCAAGATATGAGGCCCACTCCCGATACGAGCCC 60
QY 353 CACCTGAGATATGAGGCCCACTGCAAGATATGAGGCCCGCTGTGAGATACAGAG 412
DB 61 CACCTGAGATATGAGGCCCACTGCAAGATATGAGGCCCGCTGTGAGATACAGAG 120
QY 413 CCTCACTGTGCGATATGAGGCCCACTGCGATATGAGGCCCACTGCGAGATATG 472
DB 121 CCTCACTGTGCGATATGAGGCCCACTGCGATATGAGGCCCACTGCGAGATATG 180
QY 473 GAGGCCCACTGATATGAGGCCCACTGCTTGTATATGAAACCCCACTCTGGAT 532
DB 181 GAGGCCCACTGATATGAGGCCCACTGCTTGTATATGAAACCCCACTCTGGAT 240
QY 533 ATGAGGCCCACTCTGATATGAGGCCCACTGCGAATAATGAGGCCCGCTGCGG 592
DB 241 ATGAGGCCCACTCTGATATGAGGCCCACTGCGAATAATGAGGCCCGCTGCGG 300

```

```

QY 593 GATACAGAGCTTCACCTGTGATCAGAGCCAGGCTCAGATATACAGAGCCAGG 652
DB 301 GATACAGAGCTTCACCTGTGATCAGAGCCAGGCTCAGATATACAGAGCCAGG 360
QY 653 CTCCTGAAAACGAGGCTTCTTCTGCTGCTGCTCTTCTTCTGAGTTCATTCT 705
DB 361 CTCCTGAAAACGAGGCTTCTTCTGCTGCTGCTCTTCTTCTGAGTTCATTCT 413
RESULT 7
AAI48070
ID AAI48070 standard; DNA; 436 BP.
AC
XX AAI48070;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #16756 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN M0200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001MO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-48897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID NO 16756; 654bp; English.
XX
XX
CC The present invention relates to single exon nucleic acid probes (SNP).
CC The present sequence is one such probe. The probes are useful for
CC predicting a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
Query Match 58.4%; Score 411.4; DB 22; Length 436;
Best Local Similarity 99.8%; Pred. No. 9.3e-113;
Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 293 TTATTGCTATGAGGCCCCCACTGCAAGATATGAGGCCCACTCCCGATACGAGCCC 352
DB 1 TTATTGCTATGAGGCCCCCACTGCAAGATATGAGGCCCACTCCCGATACGAGCCC 60
QY 353 CACCTGAGATATGAGGCCCACTGCAAGATATGAGGCCCGCTGTGAGATACAGAG 412
DB 61 CACCTGAGATATGAGGCCCACTGCAAGATATGAGGCCCGCTGTGAGATACAGAG 120
QY 413 CCTCACTGTGCGATATGAGGCCCACTGCGATATGAGGCCCACTGCGAGATATG 472
DB 121 CCTCACTGTGCGATATGAGGCCCACTGCGATATGAGGCCCACTGCGAGATATG 180

```

QY 473 GAGCCACCTCTAGATATGAGAGCCCACTCTTGTATATGAAACCCACCTCTCGAT 532
 DB 181 GAGCCACCTCTAGATATGAGAGCCCACTCTTGTATATGAAACCCACCTCTCGAT 240
 QY 533 ATGAGAGCCCACTCTCGATATGAGAGCCCACTCTGAGAAATGAGAGCCCGCTGCG 592
 DB 241 ATGAGAGCCCACTCTCGATATGAGAGCCCACTCTGAGAAATGAGAGCCCGCTGCG 300
 QY 593 GATACAGAGCTCTACCTCTGATATGAGAGCCCGCTCTAGAAATTAAGAGAGCCAG 652
 DB 301 GATACAGAGCTCTACCTCTGATATGAGAGCCCGCTCTAGAAATTAAGAGAGCCAG 360
 QY 653 CTCTGAAAACGAGGCTTCTCTCCCTGCTCTCTCTGATGCTCAATTC 705
 DB 361 CTCTGAAAACGAGGCTTCTCTCCCTGCTCTCTCTGATGCTCAATTC 413
 RESULT 8
 ABS16034
 ID ABS16034 standard; DNA; 436 BP.
 XX
 AC ABS16034;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human genome-derived single exon probe ORF from lung SEQ ID No 16025.
 XX
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary hemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease; open reading frame; ORF.
 XX
 OS Homo sapiens.
 XX
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001MO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2002-114183/15.
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX
 PS Claim 4; SEQ ID No 16025; 634bp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridize at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a

CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridization of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridization to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC hemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karsagen syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPRO at
 CC ftp.wipro.int/pub/published_pct_sequences.
 XX
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 58.4%; Score 411.4; DB 24; Length 436;
 Best Local Similarity 99.8%; Pred. No. 9.3e-113;
 Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 293 TTATGTCTATATGAGGCGCCCACTGAGATATGAGAGCCCACTCCGATACGAGCC 352
 DB 1 TTATGTCTATATGAGGCGCCCACTGAGATATGAGAGCCCACTCCGATACGAGCC 60
 QY 353 GACCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCGCTGAGATACAGAG 412
 DB 61 GACCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCGCTGAGATACAGAG 120
 QY 413 CCTCACTGTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATG 472
 DB 121 CCTCACTGTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATG 180
 QY 473 GAGCCACCTCTAGATATGAGAGCCCACTCTGATATGAAACCCACCTCTCGAT 532
 DB 181 GAGCCACCTCTAGATATGAGAGCCCACTCTGATATGAAACCCACCTCTCGAT 240
 QY 533 ATGAGAGCCCACTCTCGATATGAGAGCCCACTCTGAGAAATGAGAGCCCGCTGCG 592
 DB 241 ATGAGAGCCCACTCTCGATATGAGAGCCCACTCTGAGAAATGAGAGCCCGCTGCG 300
 QY 593 GATACAGAGCTCTACCTCTGATATGAGAGCCCGCTCTAGAAATTAAGAGAGCCAG 652
 DB 301 GATACAGAGCTCTACCTCTGATATGAGAGCCCGCTCTAGAAATTAAGAGAGCCAG 360
 QY 653 CTCTGAAAACGAGGCTTCTCTCCCTGCTCTCTCTGATGCTCAATTC 705
 DB 361 CTCTGAAAACGAGGCTTCTCTCCCTGCTCTCTCTGATGCTCAATTC 413
 RESULT 9
 AAS20601
 ID AAS20601 standard; cDNA; 1413 BP.
 XX
 AC AAS20601;

XX 09-APR-2002 (first entry)

DT DNA encoding bovine perinuclear theca 32 (PT32).

XX Testicular WW domain binding protein; hWWBP; perinuclear theca 32;

XX PT32; contraceptive; fertility; oocyte activation; vaccine;

XX immunosuppressive; spermogenesis; spermatocyst; tyrosine kinase; c-Yes;

XX immunosuppressive; bovine; gene; ss.

OS Bos sp.

XX Key Location/Qualifiers

FT primer_bind 30..50

FT /tag= a

FT /note= "Primer binding site for cDNA isolation. The

FT sequence differs from that of the forward primer

FT given in AAS20603"

FT CDS 36..977

FT /tag= b

FT /product= "PT32"

FT /note= "Perinuclear theca 32"

FT protein_bind 978..1001

FT /tag= c

FT /note= "Primer binding site for cDNA isolation. The

FT sequence differs from that of the reverse primer

FT given in AAS20604"

XX MO200190185-A2.

XX 29-NOV-2001.

XX 25-MAY-2001; 2001MO-CA00738.

XX 25-MAY-2000; 2000CA-2307128.

XX 25-MAY-2000; 2000US-206979P.

XX (TOOH) UNIV QUEBENS KINGSTON.

XX (UYOR-) UNIV OREGON HEALTH SCI.

XX PA

XX PI Oko R, Sutovsky P;

XX WPI; 2002-097644/13.

XX DR P-PSDB; AAU74604.

XX Isolated perinuclear theca 32 polypeptide that interacts with activated

PT tyrosine kinase c-Yes, for enhancing fertility; treating/diagnosing

PT diminished fertility and abnormal spermatogenesis and for providing

PT contraception -

XX Claim 10; Fig 2A-B; 103pp; English.

XX The invention describes an isolated perinuclear theca 32 (PT32)

CC polypeptide (I) which interacts with tyrosine kinase c-Yes. (I) is

CC useful for: enhancing fertility in a mammal; treating globozoospermy, by

CC expressing (I) in spermatozoa; inhibiting fertilization, by introducing

CC (I) or its antigenic fragment into a mammal to elicit an immune

CC response; enhancing the ability of round spermatids to activate oocytes;

CC treating or diagnosing diminished fertility and abnormal spermatogenesis;

CC in providing contraception; identifying fertility and abnormal spermatogenesis;

CC fertility-enhancing agents. The polynucleotide is useful for producing

CC (I) by recombinant techniques, as vaccine, as diagnostic reagents, and

CC for chromosome identification. An antibody against (I) is useful in

CC immunological assays, in immunosuppressive methods, to identify cells

CC expressing (I), and to purify (I) by affinity chromatography. A

CC transgenic animal is useful as an animal model for studying human

CC fertility and reproductive biology, and for screening compounds to

CC identify modulators of oocyte activation. The use of (I) prevents the

CC entry of components which are detrimental to embryonic development into

CC the oocyte during oocyte activation with crude sperm extract and avoids

CC the propagation of viruses such as HIV (human immunodeficiency virus) and

CC SIV (simian immunodeficiency virus) carried in the sperm. This sequence

CC encodes the bovine perinuclear theca 32 (PT32), described in the method

CC of the invention.

XX Sequence 1413 BP; 377 A; 363 C; 369 G; 304 T; 0 other;

SQ Query Match 54.6%; Score 385; DB 24; Length 1413;

Best Local Similarity 74.0%; Pred. No. 1.3e-104;

Matches 538; Conservative 0; Mismatches 165; Indels 24; Gaps 3;

1 ATGCCATTTGATCTGATGAGCAACCTCAGTGTGAACCAACAGATTTGCTGCAAACTTC 60

258 ATGCCGTTTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 317

61 ATTAAGGAATCTATTCAGGAGCTCATATGTTGCTGGGAAGCAAGTACTTTTAA 120

318 ATTAAGGAACCAATTCAGGAGCTCCAGGATGTTGCTGGGAAGCAAGTACTTTTAA 377

121 TTAGCTTCAGAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 180

378 TTAGCTTCAGAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 437

181 GCTGTGCTGAGGATTTTCACTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTT 240

438 GCTGTGCTGAGGATTTTCACTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTT 497

241 TATGTAAATTAATCTGAGGAAAGGAAAT--ATGTGACTCCAGAGTACCTTGTTCAG----- 292

498 TACATATTAATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 557

293 ---TATGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 348

558 CCAATTTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 617

349 GCTGCTGCTGAGGATTTTCACTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTT 408

618 ACTCACCAAGAAAGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 677

409 AGAGCTCACCTGTCGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 468

678 GAGAGCTGCTGAGGATTTTCACTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTT 737

469 TATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 528

738 TATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 797

529 GATATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 588

798 GATATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 857

589 GCGGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 648

858 GCTGGAATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 917

649 -----CAGGCTCTGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 696

918 TCTCAGAGATCTATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 977

QY 697 GTTCATT 703

DB 978 GTTCATT 984

RESULT 10

AAU04882

ID AAU04882 standard; DNA; 7099 BP.

XX AAU04882;

AC 21-NOV-2001 (first entry)

DT Human reproductive system related antigen DNA SEQ ID NO: 7570.

XX Human reproductive system related antigen; reproductive system disorder;

XX Human; gene therapy; ds.

XX Homo sapiens.
OS
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216447.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-465570/50.
DR
XX Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 7570; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention.
XX
SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;

Query Match 53.2%; Score 375.4; DB 22; Length 7099;
Best Local Similarity 99.7%; Pred. No. 2.1e-101;
Matches 376; Conservative 1; Indels 0; Gaps 0;

QY 290 CAGTTATGTCATGAGGCCCCCAGCTGAGATATGAGCCCACTCCGAGATACGAG 349
DB 6723 CAGTTATGTCATGAGGCCCCCAGCTGAGATATGAGCCCACTCCGAGATACGAG 6782
QY 350 CCCCACCTGCAGATATGAGCCCACTGAGAAATGAAAGCCGCTGTGGATACA 409
DB 6783 CCCCACCTGCAGATATGAGCCCACTGAGAAATGAAAGCCGCTGTGGATACA 6842
QY 410 GAGGCTACCTGAGATATGAGCCCACTGAGATATGAGCCCACTGAGAT 469
DB 6843 GAGGCTACCTGAGATATGAGCCCACTGAGATATGAGCCCACTGAGAT 6902
QY 470 ATGAGGCCCACTCTAGATATGAGCCCACTCTTGTATGAAACCCCACTCTCG 529
DB 6903 ATGAGGCCCACTCTAGATATGAGCCCACTCTTGTATGAAACCCCACTCTCG 6962
QY 530 GATATGAGGCCCACTCTGATATGAGCCCACTGAGAAATGAAAGCCGCTG 589
DB 6963 GATATGAGGCCCACTCTGATATGAGCCCACTGAGAAATGAAAGCCGCTG 7022
QY 590 CGGATACAGAGCTTCACTGCTGATACAGAGCCAGGCTCAGAAATCTACAGAGCCC 649
DB 7023 CGGATACAGAGCTTCACTGCTGATACAGAGCCAGGCTCAGAAATCTACAGAGCCC 7082
QY 650 AGGCTCCTGAAAGAG 666
DB 7083 AGGCTCCTGAAAGAG 7099

RESULT 11
ID ABL97776 standard; DNA; 7099 BP.
XX
XX ABL97776;
XX
XX 21-JUN-2002 (first entry)
XX
XX Human testicular antigen encoding DNA fragment SEQ ID NO: 2428.
XX
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
XX reproductive system disorder; urinary system disorder; gene therapy;
XX cardiovascular disorder; respiratory disorder; neurological disorder;
XX gastrointestinal disease; infection; cytostatic; gene; ds.
XX
XX Homo sapiens.
XX
XX MO20015317-A2.
XX
XX PD 02-AUG-2001.
XX

PR 17-JAN-2001; 2001WO-US013329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 14-AUG-2000; 2000US-0226281.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227892.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234897.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249287.
PR 17-NOV-2000; 2000US-0249289.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Baraah SC, Ruben SM,
XX
XX WPI; 2001-483232/52.
PT Nucleic acids encoding 973 human testicular antigen polypeptides,

PT useful for preventing, diagnosing and/or treating testicular cancer -
XX
XX PS Disclosure; SEQ ID NO 2428; 766pp; English.
XX
XX The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer.
CC especially testicular cancers. The present sequence is a DNA encoding a
CC protein fragment of the invention.
XX
SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
Query Match 53.2%; Score 375.4; DB 23; Length 7099;
Best Local Similarity 99.7%; Pred. No. 2.1e-101;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 230 CAGTATATGTCATATGGGCCCCCACCCTGCAAGATATGAGCCCACTCCGAGTACGAG 349
DB 6723 CAGTATATGTCATATGAGCCCACTGCAAGATATGAGCCCACTCCGAGTACGAG 6782
QY 350 CCCCACTGCAAGATATGAGCCCACTGCAAGATATGAGCCCACTCCGAGTACGAG 409
DB 6783 CCCCACTGCAAGATATGAGCCCACTGCAAGATATGAGCCCACTCCGAGTACGAG 6842
QY 410 GAGCCTCACTGTCGATATGAGCCCACTGTCGATATGAGCCCACTCCGAGTACGAG 469
DB 6843 GAGCCTCACTGTCGATATGAGCCCACTGTCGATATGAGCCCACTCCGAGTACGAG 6902
QY 470 ATGAGCCCACTGTCGATATGAGCCCACTGTCGATATGAGCCCACTCCGAGTACGAG 529
DB 6903 ATGAGCCCACTGTCGATATGAGCCCACTGTCGATATGAGCCCACTCCGAGTACGAG 6962
QY 530 GATATGAGCCCACTGTCGATATGAGCCCACTGTCGATATGAGCCCACTCCGAGTACGAG 589
DB 6963 GATATGAGCCCACTGTCGATATGAGCCCACTGTCGATATGAGCCCACTCCGAGTACGAG 7022
QY 590 CGGATPACAGAGCTCACCTGTCGATATGAGCCCACTGTCGATATGAGCCCACTCCGAGTACGAG 649
DB 7023 CGGATPACAGAGCTCACCTGTCGATATGAGCCCACTGTCGATATGAGCCCACTCCGAGTACGAG 7082
QY 650 AGGCTCTGTAACGAG 666
DB 7083 AGGCTCTGTAACGAG 7099
RESULT 12
AAS70582
ID AAS70582 standard; cDNA; 894 BP.
XX
XX AAS70582;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX DNA encoding novel human diagnostic protein #6386.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW Food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX

PI Dmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG06395.
 XX
 PT New isolated polypeptide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 1, SEQ ID NO 6386, 103pp, English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 894 BP; 204 A; 221 C; 213 G; 256 T; 0 other;
 XX
 Query Match 24.6%; Score 173.6; DB 23; Length 894;
 Best Local Similarity 86.8%; Pred. No. 1.7e-41;
 Matches 191; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
 QY 448 TACGAGCCCACTGAGATATGAGCCCACTCTAGATATGAGCCCACTCTT 507
 DB 165 TATGAAGCCCGCTGTGGATACAGAGCCCTCAGTATGAGCCCACTCTT 224
 QY 508 GGATATGAGAACCCCACTCTGGAATGAGAGCCCACTCTGGAATGAGAGCCCACT 567
 DB 225 GGATATGAGAACCCCACTCTGGAATGAGAGCCCACTCTGGAATGAGAGCCCACT 284
 QY 568 GCAGGAATGAGAGCCCGCTGCGGATACAGAGCTTACCTGCTGATACAGAGCCAG 627
 DB 285 GCAGGAATGAGAGCCCGCTGCGGATACAGAGCTTACCTGCTGATACAGAGCCAG 344
 QY 628 CCTCAGGAATCTACAGAGCCCACTCTGAAAAAGAG 667
 DB 345 CCTCAGGAATCTACAGGAAGAGTGTGTCCCAAGAGCTTGG 384
 XX
 RESULT 13
 ID ABA55246
 ID ABA55246 standard; DNA; 471 BP.
 XX
 AC ABA55246;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human foetal liver single exon nucleic acid probe #3551.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.

PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632365.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI: 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 PS Claim 1, SEQ ID NO 3551; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 471 BP; 120 A; 112 C; 82 G; 157 T; 0 other;
 XX
 Query Match 18.5%; Score 130.4; DB 22; Length 471;
 Best Local Similarity 99.2%; Pred. No. 1e-28;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 290 CAGTTATTTCTATGAGGAGCCCACTGACAGATATGAGCCCACTCCGATACGAG 349
 DB 340 CAGTTATTTCTATGAGGAGCCCACTGACAGATATGAGCCCACTCCGATACGAG 399
 QY 350 CCCCACTGACAGATATGAGGAGCCCACTGAGAAATGAGAGCCCGCTGTGGATACA 409
 DB 400 CCCCACTGACAGATATGAGGAGCCCACTGAGAAATGAGAGCCCGCTGTGGATACA 459
 QY 410 GAGCTGACCTG 421
 DB 460 GAGCTGACCTG 471
 XX
 RESULT 14
 ID AAK28957
 ID AAK28957 standard; DNA; 471 BP.
 XX
 AC AAK28957;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 3514.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00668.

```
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 3514; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
XX Sequence 471 BP; 120 A; 112 C; 82 G; 157 T; 0 other;
SQ
Query Match 18.5%; Score 130.4; DB 22; Length 471;
Best Local Similarity 99.2%; Pred. No. 1e-28;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 290 CAGTTATTGTCTATGAGGCCCCCAGCTGCAAGATATGAGCCCCCACTCCCGATACGAG 349
DB 340 CAGTTATTGTCTATGAGGCCCCCAGCTGCAAGATATGAGCCCCCACTCCCGATACGAG 399
QY 350 CCCACCTGCAAGATATGAGGCCCCCAGCTGCAAGATATGAGGCCCCCGCTGTGGATACA 409
DB 400 CCCACCTGCAAGATATGAGGCCCCCAGCTGCAAGATATGAGGCCCCCGCTGTGGATACA 459
QY 410 GAGCCTCACCTG 421
DB 460 GAGCCTCACCTG 471
RESULT 15
AAI34906
ID AAI34906 strand; DNA; 471 BP.
XX
XX AAI34906;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #3592 used to measure gene expression in human placenta sample.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
```

```
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID NO 3592; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 471 BP; 120 A; 112 C; 82 G; 157 T; 0 other;
SQ
Query Match 18.5%; Score 130.4; DB 22; Length 471;
Best Local Similarity 99.2%; Pred. No. 1e-28;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 290 CAGTTATTGTCTATGAGGCCCCCAGCTGCAAGATATGAGCCCCCACTCCCGATACGAG 349
DB 340 CAGTTATTGTCTATGAGGCCCCCAGCTGCAAGATATGAGCCCCCACTCCCGATACGAG 399
QY 350 CCCACCTGCAAGATATGAGGCCCCCAGCTGCAAGATATGAGGCCCCCGCTGTGGATACA 409
DB 400 CCCACCTGCAAGATATGAGGCCCCCAGCTGCAAGATATGAGGCCCCCGCTGTGGATACA 459
QY 410 GAGCCTCACCTG 421
DB 460 GAGCCTCACCTG 471
Search completed: December 16, 2003, 11:00:51
Job time : 189.14 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 09:55:15 (Search time 49.6677 Seconds
(without alignments)
6265.145 Million cell updates/sec

Title: US-09-864-291-11_COPY_1_705

Perfect score: 1 atgcattgatcgtgatcgcac.....cctctctcagtcattctt 705

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgm2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgm2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgm2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgm2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgm2_6/ptodata/2/ina/PCTUS.COMB.seq:*
6: /cgm2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77.2	11.0	1885	4 US-09-484-970B-87	Sequence 87, Appl
2	76	10.8	981	2 US-08-841-349-15	Sequence 15, Appl
3	54.4	7.7	243	1 US-08-182-175A-56	Sequence 56, Appl
4	54.4	7.7	243	1 US-08-474-633A-74	Sequence 74, Appl
5	54.4	7.7	243	4 US-08-823-771-74	Sequence 74, Appl
6	54.4	7.7	243	5 PCT-US92-06412-56	Sequence 56, Appl
7	53.6	7.6	3833	1 US-08-917-320-18	Sequence 18, Appl
8	53.6	7.6	5931	5 PCT-US95-04611A-18	Sequence 18, Appl
9	53.6	7.6	5931	3 US-08-783-774-1	Sequence 1, Appl
10	53.6	7.6	5931	4 US-09-556-706B-1	Sequence 1, Appl
11	52	7.4	2188	1 US-07-865-662F-10	Sequence 10, Appl
12	52	7.4	2188	1 US-08-374-219B-10	Sequence 10, Appl
13	52	7.4	3489	2 US-08-728-323A-1	Sequence 1, Appl
14	52	7.4	3489	4 US-09-298-568-1	Sequence 1, Appl
15	52	7.4	3489	4 US-09-410-399-1	Sequence 1, Appl
16	52	7.4	32207	2 US-08-770-379-20	Sequence 20, Appl
17	52	7.4	32207	4 US-08-757-669A-20	Sequence 20, Appl
18	52	7.4	32207	4 US-09-230-371A-20	Sequence 20, Appl
19	50.2	7.1	604	4 US-09-370-838-175	Sequence 175, Appl
20	50	7.0	7785	2 US-08-276-967-1	Sequence 1, Appl
21	49.6	6.9	3222	4 US-08-714-741-39	Sequence 39, Appl
22	48.6	6.9	2633	4 US-09-146-476-51	Sequence 51, Appl
23	47.2	6.7	5661	4 US-08-938-105-2	Sequence 2, Appl
24	44.6	6.3	1562	4 US-09-620-312D-510	Sequence 510, Appl
25	44.6	6.3	2144	3 US-08-834-306-15	Sequence 15, Appl
26	44.6	6.3	2144	3 US-08-993-674A-15	Sequence 15, Appl
27	44.6	6.3	2144	4 US-09-256-976-15	Sequence 15, Appl

C	28	44.4	6.3	1235	2 US-08-557-309B-53	Sequence 53, Appl
	29	43.2	6.1	187	1 US-08-182-175A-90	Sequence 90, Appl
	30	43.2	6.1	187	1 US-08-474-633A-78	Sequence 78, Appl
	31	43.2	6.1	187	4 US-08-823-771-78	Sequence 78, Appl
	32	43.2	6.1	187	5 PCT-US92-06412-90	Sequence 90, Appl
	33	43	6.1	8310	3 US-08-870-126-11	Sequence 11, Appl
	34	43	6.1	8310	4 US-09-445-247-11	Sequence 11, Appl
	35	43	6.1	14985	1 US-08-652-972A-6	Sequence 6, Appl
	36	43	6.1	14985	5 PCT-US96-06231A-6	Sequence 6, Appl
	37	42.4	6.0	981	2 US-08-841-349-15	Sequence 15, Appl
	38	42.2	6.0	1251	4 US-09-252-991A-444	Sequence 44, Appl
C	39	42.2	6.0	1362	4 US-09-252-991A-417	Sequence 417, Appl
	40	42	6.0	1995	1 US-08-425-069-3	Sequence 3, Appl
	41	42	6.0	1995	2 US-08-317-844B-3	Sequence 3, Appl
	42	42	6.0	6314	4 US-09-620-312D-98	Sequence 98, Appl
	43	41	5.8	1562	4 US-09-620-312D-510	Sequence 510, Appl
	44	40.6	5.8	1849	2 US-08-665-926-7	Sequence 7, Appl
	45	40.6	5.8	1849	3 US-08-740-223A-9	Sequence 9, Appl

ALIGNMENTS

```
RESULT 1
US-09-484-970B-87
Sequence 87, Application US/09484970B
Patent No. 6426186
GENERAL INFORMATION:
APPLICANT: Jones, Karen A.
APPLICANT: Volkmut, Wayne
APPLICANT: Walker, Michael G.
TITLE OR INVENTION: BONE REMODELING GENES
FILB REFERENCE: PB-0014 US
CURRENT APPLICATION NUMBER: US/09/484,970B
CURRENT FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 172
SOFTWARE: PERL Program
SEQ ID NO 87
LENGTH: 1885
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6426186 232968.10CB1
US-09-484-970B-87

Query Match 11.0%; Score 77.2; DB 4; Length 1885;
Best Local Similarity 62.4%; Pred. No. 8.5e-14;
Matches 121; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 ATGCCATTGATGATGATGACGAACCTCACTGTTGAACACAGTATTGCTGCAAACTTC 60
|||||
DB 274 ATCCATTATTCATCATGAAAGACTGTGAGATCAACAGCCCTATTTGGTGAACACTAC 333
|||||

QY 61 ATTAAGGGAATCTTTCAGGAGCTCATATGCGCTGCGGAGAGACAGACTACTTTTAA 120
|||||
DB 334 ATCAAGGGAACACTGAAAGCGGAGGAGGCTGGAGAGCTCTGCTTCTTCAAG 393
|||||

QY 121 TTAGTCTTCAGAAATGAGATGCAATTTGATGCAATTTGATGCAATTTGATGCAATTTG 180
|||||
DB 394 TTGACTTTCACGCGAGGAGGCGCATTTGACTTTCGACAGCGAGATCTTCAGATGCGACTTC 453
|||||

QY 181 GCTGTGCGCCGAGG 194
DB 454 CAAGCCTCCAGAGG 467

RESULT 2
US-08-841-349-15/c
Sequence 15, Application US/08841349B
Patent No. 5955594
GENERAL INFORMATION:
APPLICANT: MISHRA, LOPA
```

TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT. . .
FILE REFERENCE: XX/PO4470050
CURRENT APPLICATION NUMBER: US/08/841,349B
CURRENT FILING DATE: 1997-04-30
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO: 15
LENGTH: 981
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: For all n's in this sequence, n=(a or g or c or t)
US-08-841-349-15

Query Match 10.8%; Score 76; DB 2; Length 981;
Best Local Similarity 53.2%; Pred. No. 1.4e-13;
Matches 160; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 314 CTGCGAGATATGAGGCCCACTCCCGGATACGAGAGCCCACTCGAGATATGAGAGCC 373
DB 948 CTCGAGGCCATGTAGCCAGCTCTCCAGGCCATGTAGCCAGCTCTCCAGGCCATGTAGCCA 889
QY 374 AACCCGTAGAAATGAAAGCCCGCTGTGGGATACAGAGCTCACCTGTGCATATGAGAG 433
DB 888 GCTTCCAGGCCATGTAGCCAGCTCTCCAGGCCATGTAGCCAGCTCTCCAGGCCATGTAG 829
QY 434 CCCACCTTGTGATACGAGCCCACTCGAGATATGAGAGCCCACTCGATATGATAG 493
DB 828 CCAGCTCTCAGGCCATGTAGCCAGCTCTCCAGGCCATGTAGCCAGCTCTCCAGGCCATGTAG 769
QY 494 GAGCCCACTTGTGATATGAGAGCCCACTCGAGATATGAGAGCCCACTCTGTGAT 553
DB 768 TGGCGAGATCTCAGAGACATGTGTGAGAGCTCTCCAGACATGTAGCCAGCTCTCCAGGCC 709
QY 554 ATGAGAGCCCACTCGAGAGAAATGAGAGCCCGCTGGGGATACAGAGCTCACCTGTG 613
DB 708 ATGTGGGAGCTCTCAGAGCCATGTAGCCAGCTCTCCAGGCCATTTGGAGCTCTCCAG 649
QY 614 G 614
DB 648 G 648

RESULT 3
US-08-182-175A-56
Sequence 56, Application US/08182175A
Patent No. 5559223
GENERAL INFORMATION:
APPLICANT: Saverio Carl Palco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E.I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,175A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:

NAME: Linda Axamechy Ployd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DH5 alpha
IMMEDIATE SOURCE:
CLONE: 2-9
FEATURE:
NAME/KEY: CDS
LOCATION: 2..235
OTHER INFORMATION: /function= "synthetic storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "asp"
OTHER INFORMATION: /standard_name= "7.7.7.7.7.8.9.8.9.5"
US-08-182-175A-56

Query Match 7.7%; Score 54.4; DB 1; Length 243;
Best Local Similarity 53.2%; Pred. No. 2.9e-07;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 302 ATGGAGCCCACTCGAGATATGAGCCCACTCCCGATACGAGAGCCCACTCGAG 361
DB 2 ATGAGAGAGAAAGCTGAAGCGATGAGAGAGAAAGCTGAAGCGATGAGAGAGAAAGCTGAAG 61
QY 362 GATATGAGAGCCCACTCGAGATATGAGAGCCCGCTGTGGATACAGAGCTCACCTG 421
DB 62 GCGATGAGAGAGAAAGCTGAAGCGATGAGAGAGAAAGCTGAAGCGATGAGAGAGAAAGCTG 121
QY 422 TGGATATGAGAGCCCACTCTTGGATACGAGAGCCCACTCGAGATATGAGAGCCCACTG 481
DB 122 AAGCGATGAGAGAAAGCTTAAGAAAGATGAGAAAGCTTAAGAAAGATGAGATGAGAGAGAA 181
QY 482 CTCTAGATATGAGAGCCCACTCTTGGATATGAGAA 517
DB 182 CTCAGAAAGATGAGAGAAAGCTTAAGATGATGAGAA 217

RESULT 4
US-08-474-633A-74
Sequence 74, Application US/08474633A
Patent No. 5773691
GENERAL INFORMATION:
APPLICANT: E. I. DU PONT DE NEMOURS AND
APPLICANT: COMPANY
TITLE OF INVENTION: CHIMERIC GENES AND
TITLE OF INVENTION: METHODS FOR INCREASING
TITLE OF INVENTION: INCREASING THE LYSINE
TITLE OF INVENTION: AND THREONINE CONTENT
TITLE OF INVENTION: OF THE SEEDS OF PLANTS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS
ADDRESSEE: AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,633A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGEL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; STRAIN: E. coli
; CELL TYPE: DH5 alpha
; IMMEDIATE SOURCE:
; CLONE: 2-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..235
; OTHER INFORMATION: /function= "synthetic
; OTHER INFORMATION: storage protein
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /gene= "gsp"
; OTHER INFORMATION: /standard_name=
; OTHER INFORMATION: "7.7.7.7.7.8.9.8.9.5"
; US-08-474-633A-74

Query Match 7.7%; Score 54.4; DB 1; Length 243;
Best Local Similarity 53.2%; Pred. No. 2.9e-07;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 302 ATGGGGCCCCACCTGCGATATGAGCCCCACCTCCCGATACGAGCCCCACCTGCGAG 361
DB 2 ATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAGCGATGAGAGAGAGAGCTGAAG 61
QY 362 GATATGAGAGCCCAACCGTAGAATGAGAGCCCGCTGTGGATACAGAGCTTACCTG 421
DB 62 GCATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGCTG 121
QY 422 TGCATATGAGAGCCCACTTGTGATACGAGAGCCCACTGCGAGATATGAGAGCCCACT 481
DB 122 AAGCGATGAGAGAGAGCTTAAGAGATGAGAGAGAGAGCTGAAGATGAGAGAGAGAA 181
QY 482 CTCTAGATATGAGAGCCCACTTGTGATATGAA 517
DB 182 CTCAGAAAGATGAGAGAGAGCTTAATGATGAA 217

RESULT 5
US-08-823-771-74
; Sequence 74, Application US/08823771
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
; COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND
; METHODS FOR INCREASING
; INCREASING THE LYSINE
; AND TREONINE CONTENT
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; AND COMPANY
```

```

; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,771
; FILING DATE: 24-Mar-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/474,633
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGEL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; STRAIN: E. coli
; CELL TYPE: DH5 alpha
; IMMEDIATE SOURCE:
; CLONE: 2-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..235
; OTHER INFORMATION: /function= "synthetic
; OTHER INFORMATION: storage protein
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /gene= "gsp"
; OTHER INFORMATION: /standard_name=
; OTHER INFORMATION: "7.7.7.7.7.8.9.8.9.5"
; SEQUENCE DESCRIPTION: SEQ ID NO: 74:
; US-08-823-771-74

Query Match 7.7%; Score 54.4; DB 4; Length 243;
Best Local Similarity 53.2%; Pred. No. 2.9e-07;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 302 ATGGGGCCCCACCTGCGATATGAGCCCCACCTCCCGATACGAGCCCCACCTGCGAG 361
DB 2 ATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAGCGATGAGAGAGAGAGCTGAAG 61
QY 362 GATATGAGAGCCCAACCGTAGAATGAGAGCCCGCTGTGGATACAGAGCTTACCTG 421
DB 62 GCATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGCTG 121
QY 422 TGCATATGAGAGCCCACTTGTGATACGAGAGCCCACTGCGAGATATGAGAGCCCACT 481
DB 122 AAGCGATGAGAGAGAGCTTAAGAGATGAGAGAGAGAGCTGAAGATGAGAGAGAGAA 181
QY 482 CTCTAGATATGAGAGCCCACTTGTGATATGAA 517
DB 182 CTCAGAAAGATGAGAGAGAGCTTAATGATGAA 217

RESULT 6
PCT-US92-06412-56
; Sequence 56, Application PC/TUS9206412
```


RESULT 8

PCT-US95-04611A-18
; Sequence 18, Application PC/RTUS9504611A
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
; TITLE OF INVENTION: Non Splicing Variants of gp350/220
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Cooley Godward Castro Huddleston & Tatum
; STREET: 5 Palo Alto Square
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04611A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/229,291
; FILING DATE: April 18, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luann Cseri
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR-003/0005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-843-5163
; TELEFAX: 415-857-0663
; TELEX: 380816 CooleyPA
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3833 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1014..3734
; PCT-US95-04611A-18

Query Match 7.6%; Score 53.6; DB 5; Length 3833;

Best Local Similarity 50.4%; Pred. No. 2.3e-06;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 307 GCCCACCCTGAGATATGAGGCCCACTCCCGATACGAGGCCCACTGAGATAT 366
DB 2578 GCCCACCCTGAGATATGAGGCCCACTCCCGATACGAGGCCCACTGAGATAT 2637
QY 367 GAGGCCCAACCCGTAGAGAAATGAGGCCCGCTGTGAGATACAGAGCTCACTGTGCA 426
DB 2638 CTACCCCAACCCCAATGACACGAGGCCCACTTGGGAAAAACAAGTCTTACCTGAGCAG 2697
QY 427 TATGAGGCCCACTCTTGGATATGAGAGGCCCACTGAGAGATATGAGGCCCACTCTTA 486
DB 2698 TGACTACCCCAACCCCAATGACACGAGGCCCACTTGGGAAAAACAAGTCTTACCTGAG 2757
QY 487 GGATATGAGGCCCACTCTTGGATATGAGAGGCCCACTCTGAGATATGAGGCCCACTCTTA 546
DB 2758 CAGTACTACCCCAACCCCAATGACACGAGGCCCACTTGGGAAAAACAAGTCTTACCTGAG 2817
QY 547 CTCGATATGAGGCCCACTCTTGGATATGAGAGGCCCACTCTGAGATATGAGGCCCACTCTTA 546
DB 2818 CAGCAGTACTACCCCAACCCCAATGACACGAGGCCCACTTGGGAAAAACAAGTCTTACCTGAG 2837

RESULT 9

US-08-783-774-1
; Sequence 1, Application US/08783774
; Patent No. 6054130
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; TITLE OF INVENTION: NON-SPLICING VARIANTS OF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/783,774
; FILING DATE: 15-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7682-037
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-863-8864
; TELEX: 66141 PENNIR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5931 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1014..3734
; OTHER INFORMATION:
; US-08-783-774-1

Query Match 7.6%; Score 53.6; DB 3; Length 5931;

Best Local Similarity 50.4%; Pred. No. 2.9e-06;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 307 GCCCACCCTGAGATATGAGGCCCACTCCCGATACGAGGCCCACTGAGATAT 366
DB 2578 GCCCACCCTGAGATATGAGGCCCACTCCCGATACGAGGCCCACTGAGATAT 2637
QY 367 GAGGCCCAACCCGTAGAGAAATGAGGCCCGCTGTGAGATACAGAGCTCACTGTGCA 426
DB 2638 CTACCCCAACCCCAATGACACGAGGCCCACTTGGGAAAAACAAGTCTTACCTGAGCAG 2697
QY 427 TATGAGGCCCACTCTTGGATATGAGAGGCCCACTGAGAGATATGAGGCCCACTCTTA 486
DB 2698 TGACTACCCCAACCCCAATGACACGAGGCCCACTTGGGAAAAACAAGTCTTACCTGAG 2757
QY 487 GGATATGAGGCCCACTCTTGGATATGAGAGGCCCACTCTGAGATATGAGGCCCACTCTTA 546
DB 2758 CAGTACTACCCCAACCCCAATGACACGAGGCCCACTTGGGAAAAACAAGTCTTACCTGAG 2817
QY 547 CTCGATATGAGGCCCACTCTTGGATATGAGAGGCCCACTCTGAGATATGAGGCCCACTCTTA 546
DB 2818 CAGCAGTACTACCCCAACCCCAATGACACGAGGCCCACTTGGGAAAAACAAGTCTTACCTGAG 2837

RESULT 10

US-09-556-706B-1
Sequence 1, Application US/09556706B
Patent No. 6458364
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
TITLE OF INVENTION: NON SPLICING VARIANTS OF GP350/220
FILE REFERENCE: 7682-050-999
CURRENT APPLICATION NUMBER: US/09/556,706B
CURRENT FILING DATE: 2000-04-24
PRIOR APPLICATION NUMBER: 08/783,774
PRIOR FILING DATE: 1997-01-15
PRIOR APPLICATION NUMBER: 08/229,291
PRIOR FILING DATE: 1994-04-18
NUMBER OF SEQ ID NOS: 19
SOFTWARE: Patent in version 3.0
SEQ ID NO 1
LENGTH: 5931
TYPE: DNA
ORGANISM: Virus
FEATURE:
OTHER INFORMATION: gp350/220
US-09-556-706B-1

Query Match 7.6%; Score 53.6; DB 4; Length 5931;
Best Local Similarity 50.4%; Pred. No. 2.9e-06;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 307 GCCCAGCTGAGATATGAGAGCCCACTCCCGATGAGAGAGCCCACTTCAGAGATAT 366
DB 2578 GCGCCAGCCGAGAGTACTACCCCAAAATGCCAGAGCCCACTTCAGAGATAT 2637
QY 367 GAGGCCCAACCGTATGAGAGAGCCCGCTGTGGATACAGAGCTTACCTTGCA 426
DB 2638 CTAACCCCAACCAATGAGAGAGCCCACTTCGAGAGAGAGAGAGAGAGAG 2697
QY 427 TATGAGAGCCCACTTGTATGAGAGAGCCCACTTCGAGAGAGAGAGAGAGAG 486
DB 2698 TACTACCCCAACCAATGAGAGAGAGCCCACTTCGAGAGAGAGAGAGAGAG 2757
QY 487 GATATGAGAGCCCACTTGTATGAGAGAGAGCCCACTTCGAGATATGAGAGAG 546
DB 2758 CAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2817
QY 547 CTCGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 566
DB 2818 CAGCAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2837

RESULT 11
US-07-865-662F-10
Sequence 10, Application US/07865662F
Patent No. 5451670
GENERAL INFORMATION:
APPLICANT: Marcia M. Miller
TITLE OF INVENTION: Restriction Fragment Length
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSER: City of Hope
STREET: 1500 East Duarte Road
CITY: Duarte
STATE: California
COUNTRY: United States of America
ZIP: 91010-0269
COMPUTER READABLE FORM:
MEDIUM TYPE: 3M Double Density 5 1/4" diskette
COMPUTER: Wang PC
OPERATING SYSTEM: MS DOS Version 3.20
SOFTWARE: Microsoft
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/865,662F
FILING DATE: 07 April, 1992

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/688,326
FILING DATE: 22 April 1991
APPLICATION NUMBER: 07/588,922
FILING DATE: 27 September 1990
APPLICATION NUMBER: 07/210,405
FILING DATE: 23 June 1988
APPLICATION NUMBER: US 07/130,529
FILING DATE: 9 December 1987
APPLICATION NUMBER: US 07/068,176
FILING DATE: 30 June 1987
ATTORNEY/AGENT INFORMATION:
NAME: Irons, Edward S.
REGISTRATION NUMBER: 16,541
REFERENCE/DOCKET NUMBER: NO. 5451670e
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 783-6040
TELEFAX: (202) 783-6031
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2188
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: DNA
ORIGINAL SOURCE: Synthetically Prepared
IMMEDIATE SOURCE: Synthetically Prepared
US-07-865-662F-10

Query Match 7.4%; Score 52; DB 1; Length 2188;
Best Local Similarity 47.5%; Pred. No. 5.2e-06;
Matches 154; Conservative 0; Mismatches 170; Indels 0; Gaps 0;
QY 302 ATGGGAGCCCACTGAGATATGAGAGCCCACTCCGATGAGAGAGAGAGAGAG 361
DB 1315 ATGGGCAACAAGCTTAAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
QY 362 GATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
DB 1375 GAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
QY 422 TGGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 481
DB 1435 GAGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494
QY 482 CTCTAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541
DB 1495 CATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1554
QY 542 CACTCTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 601
DB 1555 CACCATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614
QY 602 CCTCAGCTGCTGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 625
DB 1615 GAGCAGCATGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1638

RESULT 12
US-08-374-219B-10
Sequence 10, Application US/08374219B
Patent No. 6218106
GENERAL INFORMATION:
APPLICANT: Miller, Marcia M.
TITLE OF INVENTION: Restriction Fragment Length Polymorphism Test
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSER: Rothwell, F199, Ernst & Kurtz
STREET: 555 Thirteenth Street, N.W.
CITY: Washington

STATE: D. C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/374,219B
FILING DATE: 18-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/865,662
FILING DATE: 07-APR-1992
APPLICATION NUMBER: 07/688,326
FILING DATE: 22-APR-1991
APPLICATION NUMBER: 07/588,922
FILING DATE: 27-SEP-1990
APPLICATION NUMBER: 07/413,301
FILING DATE: 28-SEP-1989
APPLICATION NUMBER: 07/210,405
FILING DATE: 23-JUN-1988
APPLICATION NUMBER: 07/130,529
FILING DATE: 09-DEC-1987
APPLICATION NUMBER: 07/068,176
FILING DATE: 30-JUN-1987
ATTORNEY/AGENT INFORMATION:
NAME: Walker, Barbara W
REGISTRATION NUMBER: 35,400
REFERENCE/DOCKET NUMBER: 2124-113A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 783-6040
TELEFAX: 202 783-6031
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2188 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-374-219B-10

Query Match 7.4%; Score 52; DB 3; Length 2188;
Best Local Similarity 47.5%; Pred. No. 5.2e-06;
Matches 154; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 302 ATGGGGCCCCACCTGAGAGATATGAGGCCCACTCCCGGATACGAGCCCACTGAG 361
DB 1315 ATGGGGCAACAAGCTAAAGAAATCAGAGGAACAAGAAATCGAGCTGAAGAGCACCAGAG 1374
QY 362 GATATGAGCCCAACCCGTAGAAATGAAGCCCGCTGTGAGATACAGACCTCAGCTG 421
DB 1375 GAGATGGGCAACAAGCTAAAGAAATCAGAGGAACAAGAAATCGAGCTGAAGAGCACCAT 1434
QY 422 TGCATATGAGCCCACTCTTGATACGAGCCCACTGCAAGATATGAGCCCAAC 481
DB 1435 GAGAGATGGGCAACAAGCTAAAGAAATCAGAGGAACAAGAAATCGAGCTGAAGAGC 1494
QY 482 CTATGATATGAGCCCACTCTTGATATGAAACCCCACTCTCGGATATGAGCC 541
DB 1495 CATAGGAGATGGGCAACAAGCTAAAGAAATCAGAGGAACAAGAAATCGAGCTGAAGAG 1554
QY 542 CACCTCTGATATGAGCCCACTGCAAGAAATGAAGCCCGCTGCGGATACAGAG 601
DB 1555 CACCATGAGAGATGGGCAACAAGCTAAAGAAATCAGAGGAACAAGAAATCGAGCTGAAG 1614
QY 602 CTTCACTGTGATGAGAGCCCA 625
DB 1615 GAGCACCATGAGAGATGGGCA 1638

RESULT 13

US-08-728-323A-1
Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Rusco, James J.
APPLICANT: Edelman, Isidore S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSES: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-391-0525
TELEFAX: 212-278-0400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 7.4%; Score 52; DB 2; Length 3489;
Best Local Similarity 48.1%; Pred. No. 6.7e-06;
Matches 148; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 316 GAGATATGAGAGCCCACTCCCGGATACGAGCCCACTGAGATATGAGCCCA 375
DB 2202 GAGCAGACACAGATAGCAGCAGCAGATGACAGCAGCAGATGAAACGGA 2261
QY 376 CCGTAAATGAGAGCCCGCTGTGAGATACAGAGCTCAGCTGTGATATGAGCC 435
DB 2262 GAGCAGAGAGAGAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2321
QY 436 CCACTCTTGATAGAGAGCCCACTGCAAGATATGAGAGCCCACTTATGATATG 495
DB 2322 GAGATTAGAGATCAGAGAGAGAGATTAGAGCAGAGCAGAGATTAGAGAGAG 2381
QY 496 GCCCACCTCTGATATGAGAGCCCACTCTGAGATATGAGAGCCCACTCTGATAT 555
DB 2382 GAGAGATTAGAGAGAGAGAGAGAGAGATTAGAGAGAGAGAGAGAGAGAG 2441
QY 556 GAGAGCCCACTGAGAGAAATGAAGCCCGCTGCGGATACAGAGCTCAGCTGAT 615
DB 2442 GAGCAGAGATTAGAGAGAGAGAGAGAGAGATTAGAGAGAGAGAGAGAGAG 2501
QY 616 TCAGAGAC 623

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 10:43:10 : Search time 232.016 Seconds

(without alignments)
10099.015 Million cell updates/sec

Title: US-09-864-291-11_COPY_1_705

Perfect score: 705

Sequence: 1 atgcacattcctcgtacgac.....cctcttcacgtcattcct 705

Scoring table:

Gapop 10.0, Gapext 1.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	446.2	58.4	467	11	US-09-764-891-1230 Sequence 1230, Ap
2	411.4	58.4	436	9	US-09-864-761-1233 Sequence 31561, A
3	375.4	53.2	7099	11	US-09-764-891-1570 Sequence 7570, A
4	130.4	18.5	471	9	US-09-864-761-15029 Sequence 15029, A
5	110.2	15.6	321	11	US-09-764-891-1356 Sequence 1356, Ap
6	109	15.5	866	13	US-10-029-386-22839 Sequence 22839, A
7	100	14.2	1040	13	US-10-029-386-26033 Sequence 26033, A
8	93.2	13.2	291	11	US-09-764-891-7571 Sequence 7571, Ap
9	89.4	12.7	1984	9	US-09-864-761-4562 Sequence 4562, Ap
10	88.8	12.6	541	13	US-10-029-386-12333 Sequence 12333, A
11	87.8	12.5	511	13	US-10-029-386-24981 Sequence 24981, A
12	87	12.3	543	13	US-09-864-761-21308 Sequence 21308, A
13	77.2	11.0	432	11	US-09-918-995-33536 Sequence 33536, A
14	69.2	9.8	541	13	US-10-029-386-11265 Sequence 11265, A
15	67.8	9.6	511	13	US-10-029-386-24981 Sequence 24981, A

16	64.8	9.2	593	13	US-10-027-632-290832 Sequence 290832, A
17	64.8	9.2	593	14	US-10-027-632-290832 Sequence 290832, A
18	63.2	9.0	409	10	US-09-864-761-148124 Sequence 148124, A
19	62.2	8.8	689	13	US-10-027-632-148124 Sequence 148124, A
20	62.2	8.7	509	9	US-09-864-761-25366 Sequence 25366, A
21	61.2	8.7	596	9	US-09-864-761-8648 Sequence 8648, Ap
22	61.2	8.7	596	9	US-09-864-761-8648 Sequence 8648, Ap
23	57.4	8.1	2489	11	US-09-298-5238-72 Sequence 72, Ap1
24	57.2	8.1	436	9	US-09-864-761-1561 Sequence 31561, A
25	57.2	8.1	7099	11	US-09-764-891-1570 Sequence 7570, Ap
26	54.4	7.7	243	15	US-10-023-0664-74 Sequence 74, Ap1
27	54.2	7.7	1218	9	US-09-864-761-20761 Sequence 20761, A
28	54.2	7.7	1954	9	US-09-864-761-4001 Sequence 4001, Ap
29	54.2	7.7	3135	15	US-10-175-523-151 Sequence 151, Ap
30	54.2	7.7	4768	10	US-09-842-777-9 Sequence 9, Ap11
31	54.2	7.7	5994	13	US-10-205-219-136 Sequence 136, Ap
32	53.2	7.5	4395	9	US-09-925-301-224 Sequence 224, Ap
33	53.2	7.5	8412	11	US-09-919-039-223 Sequence 223, Ap
34	52.8	7.5	2016	10	US-09-938-8428-2004 Sequence 2004, Ap
35	52.6	7.5	5327	13	US-10-240-453-64 Sequence 64, Ap1
36	52.6	7.5	5327	15	US-10-239-676-62 Sequence 62, Ap11
37	52	7.4	3489	13	US-10-294-804-1 Sequence 1, Ap11
38	50.2	7.1	604	10	US-09-738-973-175 Sequence 175, Ap
39	50.2	7.1	604	10	US-09-854-133-175 Sequence 175, Ap
40	50.2	7.1	604	15	US-10-144-6498-175 Sequence 175, Ap
41	50	7.1	895	9	US-09-864-761-23227 Sequence 23227, A
42	49.6	7.0	591	9	US-09-864-761-19727 Sequence 19727, A
43	49.6	7.0	1959	9	US-09-864-761-8946 Sequence 2946, Ap
44	49.6	7.0	3463	11	US-09-298-5238-59 Sequence 59, Ap1
45	48.8	6.9	5918	15	US-10-198-846-13875 Sequence 13875, A

ALIGNMENTS

RESULT 1

US-09-764-891-1230

Sequence 1230, Application US/09764891

Publication No. US20030077808A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

PRIOR REFERENCE: PC006

CURRENT APPLICATION NUMBER: US/09/764,891

PRIOR FILING DATE: 2001-01-17

Prior application data removed - consult PAM or file wrapper

NUMBER OF SEQ ID NOS: 10231

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1230

LENGTH: 467

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

NAME/KEY: SITE

LOCATION: (421)

OTHER INFORMATION: n equals a,t,g, or c

US-09-764-891-1230

Query Match	63.3%	Score 446.2	DB 11	Length 467
Best Local Similarity	98.5%	Pred. No. 1.4e-136		
Matches 459	Conservative 1	Mismatches 5	Indels 1	Gaps 1
Qy	129	CAGAAATGAGATGCCATTGAATTCGCGATTGATGATGAAAGCTGCTGCTGTC	188	
Db	2	CAGAAATGAGATGCCATTGAATTCGCGATTGATGATGAAAGCTGCTGCTGTC	61	
Qy	189	CCGAGATTTCCATTGAACTTAATGACTGTTAGCTCTTAAGGAATTATGTAAT	248	
Db	62	CCGAGATTTCCATTGAACTTAATGACTGTTAGCTCTTAAGGAATTATGTAAT	121	
Qy	249	TACTGGGAGGAGGATATGTCACCTCAGAGATGCTTTCAGTTATTTGATGGGGC	308	
Db	122	TACTGGGAGGAGGATATGTCACCTCAGAGATGCTTTCAGTTATTTGATGGAGC	181	

```
QY 309 CCCACTGAGATATGAGCCCCCACTCCCGATACGAGCCCCCACTGAGATATG 368
DB 182 CCCACTGAGATATGAGCCCCCACTCCCGATACGAGCCCCCACTGAGATATG 241
QY 369 AGCCCAACCCGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGGATA 428
DB 242 AGCCCAACCCGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGGATA 301
QY 429 TGGAGCCCACTCTTGATACGAGAGCCCACTGAGATATGAGCCCCCACTCTAG 488
DB 302 TGGAGCCCACTCTTGATACGAGAGCCCACTGAGATATGAGCCCCCACTCTAG 361
QY 489 ATATGAGCCCC-ACCTCTTGATATGAAACCCCACTCTGAGATATGAGCCCCCACTC 547
DB 362 ATATGAGCCCCCACTCTTGATATGAAACCCCACTCTGAGATATGAGCCCCCACTC 421
QY 548 TGGAGATGAGAGCCCACTGAGAGAAATGAGAGCCGCTGTGGAG 593
DB 422 TGGATATGAGAGCCCACTGAGAGAAATGAGAGCCGCTGTGGAG 467
```

RESULT 2

```
US-09-864-761-31561
Sequence 31561, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
```

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

```
SEQ ID NO 31561
LENGTH: 436
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO 299716.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
OTHER INFORMATION: EST HUMAN HIT: AW169980.1, EVALU8 8.00e-11
OTHER INFORMATION: SWISSPROT HIT: P18616, EVALU8 3.00e-13
US-09-864-761-31561
```

```
Query Match 58.4%; Score 411.4; DB 9; Length 436;
Best Local Similarity 99.8%; Pred. No. 4.3e-125;
Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 293 TTATTGTCTATGAGGCCCCCACTGAGATATGAGCCCCCACTCCGATACGAGCCC 352
DB 1 TTATTGTCTATGAGGCCCCCACTGAGATATGAGCCCCCACTCCGATACGAGCCC 60
QY 353 CACCTGAGATATGAGGCCCCCACTGAGAAATGAGGCGCTGTGGATACAGAG 412
DB 61 CACCTGAGATATGAGGCCCCCACTGAGAAATGAGGCGCTGTGGATACAGAG 120
QY 413 CCTCACTGTGTGATGAGAGCCCACTCTTGATATGAGAGCCCACTGAGATATG 472
DB 121 CCTCACTGTGTGATGAGAGCCCACTCTTGATATGAGAGCCCACTGAGATATG 180
QY 473 GAGCCCCCACTCTAGATATGAGAGCCCACTCTTGATATGAGAAACCCCACTCTGAT 532
DB 181 GAGCCCCCACTCTAGATATGAGAGCCCACTCTTGATATGAGAAACCCCACTCTGAT 240
QY 533 ATGAGGCCCCCACTCTGATATGAGAGCCCACTGAGAAATGAGGCGCTGTGGAG 592
DB 241 ATGAGGCCCCCACTCTGATATGAGAGCCCACTGAGAAATGAGGCGCTGTGGAG 300
QY 593 GATACAGAGCTCACTGATGAGAGAGCCCACTGAGATATGAGAGCCCACTGAGAG 652
DB 301 GATACAGAGCTCACTGATGAGAGAGCCCACTGAGATATGAGAGCCCACTGAGAG 360
QY 653 CTCCTGAAAACGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
DB 361 CTCCTGAAAACGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 413
```

RESULT 3

```
US-09-764-891-7570
Sequence 7570, Application US/09764891
Publication No. US2003007808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 7570
LENGTH: 7099
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-891-7570
```

```
Query Match 53.2%; Score 375.4; DB 11; Length 7099;
Best Local Similarity 99.7%; Pred. No. 1.7e-112;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 290 CAGTTATGCTATAGGAGCCCACTGAGATATGAGCCCCCACTCCGATACGAG 349
|||||
```

Db 6723 CAGTTATTGCTATGAGAGCCCACTGAGGATATGAGCCCACTCCCGGATACGAG 6782
Qy 350 CCCCACTGAGATATGAGAGCCCAACCCGTAAGAAATGAGAGCCCGCTGTGGATACA 409
Db 6783 CCCCACTGAGATATGAGAGCCCAACCCGTAAGAAATGAGAGCCCGCTGTGGATACA 6842
Qy 410 GAGCTCACTGTGCGATATGAGAGCCCACTCTTGTGATGAGAGCCCACTGTGAGAT 469
Db 6843 GAGCTCACTGTGCGATATGAGAGCCCACTCTTGTGATGAGAGCCCACTGTGAGAT 6902
Qy 470 ATGAGAGCCCACTCTGATATGAGAGCCCACTCTTGTGATGAGAGCCCACTCTCTG 529
Db 6903 ATGAGAGCCCACTCTGATATGAGAGCCCACTCTTGTGATGAGAGCCCACTCTCTG 6962
Qy 530 GATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATGAGAGAGCCCGCTG 589
Db 6963 GATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATGAGAGAGCCCGCTG 7022
Qy 590 CGGATACAGAGCTCTGATATGAGAGCCCACTCTGATGAGAGCCCACTCTGATGAGAGCC 649
Db 7023 CGGATACAGAGCTCTGATATGAGAGCCCACTCTGATGAGAGCCCACTCTGATGAGAGCC 7082
Qy 650 AGGCTCTGAAAAAGAG 666
Db 7083 AGGCTCTGAAAAAGAG 7099
RESULT 4
US-09-864-761-15029
; Sequence 15029, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DRIVEN SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aesomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15029
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO Z99716.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
US-09-864-761-15029

Query Match 18.5%; Score 130.4; DB 9; Length 471;
Best Local Similarity 99.2%; Pred. No. 2.8e-32;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 290 CAGTTATTGCTATGAGAGCCCACTGAGGATATGAGAGCCCACTCCCGATACGAG 349
Db 340 CAGTTATTGCTATGAGAGCCCACTGAGGATATGAGAGCCCACTCCCGATACGAG 399
Qy 350 CCCCACTGAGATATGAGAGCCCAACCCGTAAGAAATGAGAGCCCGCTGTGGATACA 409
Db 400 CCCCACTGAGATATGAGAGCCCAACCCGTAAGAAATGAGAGCCCGCTGTGGATACA 459
Qy 410 GAGCTCACTGT 421
Db 460 GAGCTCACTGT 471

RESULT 5
US-09-764-891-1356
; Sequence 1356, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1356
; LENGTH: 321
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (316)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (321)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1356

Query Match 15.6%; Score 110.2; DB 11; Length 321;
Best Local Similarity 96.8%; Pred. No. 1e-25;
Matches 120; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Qy 1 ATGCACTTATGATGAGAGAGCACTGATGAGAGCAAGCAAGATTTGCTGAACTTC 60
Db 195 ATGCACTTATGATGAGAGAGCACTGATGAGAGCAAGCAAGATTTGCTGAACTTC 254
Qy 61 ATTAAGGAAGTATTCAGAGAGCTCCATATGATGATGAGAGCAAGCAAGTATTTAA 119
Db 255 ATTAAGGAAGTATTCAGAGAGCTCCATATGATGATGAGAGCAAGCAAGTATTTAA 314

Qy 120 ATTA 123
Db 315 ANTA 318

RESULT 6

US-10-029-386-22839/c
; Sequence 22839, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AEMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 22839
LENGTH: 866
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005630.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
OTHER INFORMATION: SWISSPROT HIT: Q08379, EVALUOR 4.00e-10
OTHER INFORMATION: NT HIT: g16161055, EVALUOR 4.00e-95
OTHER INFORMATION: EST_HUMAN HIT: BF828725.1, EVALUOR 0.00e+00
US-10-029-386-22839

Query Match 15.5%; Score 109; DB 13; Length 866;
Best Local Similarity 54.3%; Pred. No. 4.6e-25;
Matches 220; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

Qy 297 TGTCTATGAGGAGCCCACTGACAGATATGAGGCCCACTCCCGATATGAGGCCCACTCC 356
Db 577 TGTTCACACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 518
Qy 357 TGCAGATATGAGCCCACTGAGGAAATGAGGCCCGCTGTGGATACAGAGCTTC 416
Db 517 TCTGTTCACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 458
Qy 417 ACCTGTGCAATGAGAGCCCACTCTGTGATACAGAGGCCCACTCTCGATATGAG 476
Db 457 TCTCTCTGTTCACAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 398
Qy 477 CCCACCTCTGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTCGATATGAG 536
Db 397 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 338
Qy 537 AGCCCACTCTGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTCGATATGAG 596
Db 337 AGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 278
Qy 597 CAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 656
Db 277 CACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 218
Qy 657 TGAAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 701
Db 217 TCATGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 173

RESULT 7

US-10-029-386-26033/c
; Sequence 26033, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AEMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 26033
LENGTH: 1040
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005630.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
OTHER INFORMATION: NT HIT: g16161055, EVALUOR 2.00e-97
OTHER INFORMATION: EST_HUMAN HIT: BF828725.1, EVALUOR 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: Q08379, EVALUOR 6.00e-10
US-10-029-386-26033

Query Match 14.2%; Score 100; DB 13; Length 1040;
Best Local Similarity 53.3%; Pred. No. 4.8e-22;
Matches 211; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

Qy 306 GGCCTCACTGAGATATGAGAGCCCACTCCCGATATGAGAGCCCACTGAGATATGAG 365
Db 568 GGCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 509
Qy 366 TGCAGAGAGCCCACTGAGGAAATGAGGCCCGCTGTGGATATGAGAGCTTCACCTGTCG 425
Db 508 CACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 449
Qy 426 ATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTGAGATATGAGAGCCCACTCT 485
Db 448 TCAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 389
Qy 486 AGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTGAGATATGAGAGCCCACT 545
Db 388 TGTTCACACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 329
Qy 546 TCTCGATATGAGAGCCCACTGACAGAAATGAGGCCCGCTGTGGATATGAGAGCTTC 605
Db 328 TCTGTTCACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 269
Qy 606 ACCTGTGATGAGAGAGCCCACTCTGAGATATGAGAGCCCACTCTGAGATATGAGAG 665
Db 268 TCTCATGATGAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 209
Qy 666 GGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 701
Db 208 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 173

RESULT 8

US-09-764-891-7571
; Sequence 7571, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7571
LENGTH: 291
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-891-7571

Query Match 13.2%; Score 93.2; DB 11; Length 291;
Best Local Similarity 96.9%; Pred. No. 4.1e-20;
Matches 95; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGCCATTGATCGATGACCACTCTGTTGAACACGATTTGCTGCAAACTTC 60
DB 61 ATGCCATTGATCGATGACCACTCTGTTGAACACGATTTGCTGCAAACTTC 120

QY 61 ATTAAGGAACTATTCAGGCGCTCCATATGTGCTG 98
DB 121 ATTAAGGAACTATTCAGGCGCTCCATATGTGTAAGTG 158

RESULT 9
US-09-864-761-4562/c
; Sequence 4562, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4562
; LENGTH: 1984
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009178.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
US-09-864-761-4562

Query Match 12.7%; Score 89.4; DB 9; Length 1984;
Best Local Similarity 51.6%; Pred. No. 2.2e-18;
Matches 204; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

QY 300 CTATGGGGCCCACTGATGATGAGCCCACTCCCGATGACGAGCCCACTGCG 359
DB 872 CTCGGAGAGACTCTCCAGTTCGGAGAGACTCTCCAGCTCCGAGAGACACTCC 813
QY 360 AGATATGAGAGCCCACTGAGAAATGAGAGCCGCTGTGGATATGAGAGCTCAC 419
DB 812 CAGTTCCGAGAGACTCTCCAGCTCCGAGAGACTCTCCAGTTCCGAGAGACTCC 753
QY 420 TGTGCAATATGAGAGCCCACTGATGATGAGAGCCCACTGCAATATGAGAGCC 479
DB 752 TCCAGCTCCGAGAGAGACTCTCAAGCTCTGAGAGAGACTCTCCAGTTCTGAGAGAC 693
QY 480 ACTCTGATATGAGAGCCCACTGATGATGAGAGCCCACTCTGATATGAGAGC 539
DB 692 ACCCTCCAGTTCGGAGAGACTCTCCAGTTCGGAGAGACTCTCCAGTTCCGAGAG 633
QY 540 CCCACCTCTGATATGAGAGCCCACTGAGAGAAATGAGAGCCGCTGCGATACAG 599
DB 632 GACTCTCCAGTTCGGAGAGACTCTCCAGTTCGGAGAGACTCTCCAGCTCCGAG 573
QY 600 AGCTCACTGCTGATGATGAGAGCCCACTGAGAAATATGAGAGCCCACTGAGCTCTGA 659
DB 572 AGAGACTCTCCAGTTCGGAGAGACTCTCCAGTTCGGAGAGACTCTCTCTCATTC 513
QY 660 AAGCAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 694
DB 512 CAGAGAGACTCTCCAGTTCGGAGAGACTCTCTCTCTCTCTCTCTCTCTCTCTCT 478

RESULT 10
US-10-029-386-12333/c
; Sequence 12333, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AECM1CA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 3428
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12333
; LENGTH: 541
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005630.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: NT HIT: g113651705, EVALUATE 2.00e-06
; OTHER INFORMATION: EST_HUMAN HIT: BF000754.1, EVALUATE 0.00e+00
US-10-029-386-12333

Query Match 12.6%; Score 88.8; DB 13; Length 541;
Best Local Similarity 55.5%; Pred. No. 1.6e-18;
Matches 171; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 306 GGCCCACTGAGATATGAGCCCACTCCCGATAGAGAGCCCACTGAGATA 365
DB 314 GGAGAGTTCTCTCTTCAACAGAGCTCTCTCTCTTCAAGAGCTCTCTCTTCA 255
QY 366 TGAAGCCCAACCGTAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425
DB 254 CACAGAGCTCTCTCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 195
QY 426 ATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485
DB 194 TCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 135
QY 486 AGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 545
DB 134 TGTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 75
QY 546 TCTCGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605
DB 74 TCTGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15
QY 606 ACCTGCTG 613
DB 14 TCTCATG 7

RESULT 11
US-10-029-386-24981
Sequence 24981, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AROMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 24981
LENGTH: 511
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO Z97205.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
US-10-029-386-24981

Query Match 12.5%; Score 87.8; DB 13; Length 511;
Best Local Similarity 52.3%; Pred. No. 3,4e-18;
Matches 194; Conservative 0; Mismatches 177; Indels 0; Gaps 0;
QY 283 CCTTTGCTGATATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 342
DB 138 CTTCTCTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 197
QY 343 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 402
DB 198 TCCCTCTCTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 257
QY 403 GATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 462
DB 258 GAGCTCTCTCTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 317
QY 463 GAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 522
DB 318 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 377
QY 523 CTTCTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 582
DB 378 CTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 437

QY 583 CCGCTGAGAGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 642
DB 438 CCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 497
QY 643 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 653
DB 498 CTTCTCATGAG 508

RESULT 12
US-09-864-761-21308/c
Sequence 21308, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aromica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 21308
LENGTH: 543
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL009178.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
OTHER INFORMATION: EST HUMAN HIT: AM382320.1, EVALU8 8.80e-02
OTHER INFORMATION: SWISSPROT HIT: Q05904, EVALU8 7.00e-12
US-09-864-761-21308

Query Match 12.3%; Score 87; DB 9; Length 543;
Best Local Similarity 54.1%; Pred. No. 6.5e-18;
Matches 177; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 305 GGGCCCACTGACAGATATGAGCCCACTCCGGATAGAGAGCCCACTGAGAGT 364
DB 403 GGGAACTCTCCCTCCAGTTCCGAGAGACCTTCCAGCTCCGAGAGACTCTCCACT 344
QY 365 ATGAGACCAACCCCTAGAAATGAGGCCCGCTGTGGATACAGACCTCACTGTGC 424
DB 343 CCGAGAGACTCTCTCCAGTTCCGAGAGACTCTCCAGCTCCAGAGACACTTCCA 284
QY 425 GATATGAGCCCACTCTTGGATACGAGCCCACTGAGATATGAGCCCACTCTC 484
DB 283 GTTCCGAGAGACTCTCTCCAGCTCCGAGAGACTCTCCAGTTCCGAGAGACTCTC 224
QY 485 TAGATATGAGCCCACTCTTGGATATGAGACCCCACTCTGGATATGAGCCCACT 544
DB 223 CCACTCCGAGAGACTCTCTCAAACTCTGAGAGACTCTCTCCAGTTCTGAGAGAC 164
QY 545 CTCTCGATATGAGCCCACTGAGAGAAATGAGGCCCGCTGCGGATACAGAGCT 604
DB 163 CTCCAGTCTCCGAGAGACTCTCTCCAGTTCTGAGAGACTCTCTCCAGTTCCGAGAG 104
QY 605 CACCTGCTGATCAGAGAGCCAGGCTTC 631
DB 103 CTCTCCAGTTCCGAGAGACTCTCTC 77

RESULT 13

US-09-918-995-33536
Sequence 33536, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 33536
LENGTH: 432
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(432)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-33536

Query Match 11.0%; Score 77.2; DB 11; Length 432;
Best Local Similarity 62.4%; Pred. No. 9.8e-15;
Matches 121; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 ATGCAATTGATCTGATACCAACCTCACTTTGAACAACAGATATTTGCTGAACTTC 60
DB 231 ATGCAATTGATCTGATACCAACCTCACTTTGAACAACAGATATTTGCTGAACTTC 290
QY 61 ATTAAGGAACATATTCAGAGAGCTCCATATGCTGCTGGAGAGCAAGCTACTTTAA 120
DB 291 ATCAAGGAACATATTCAGAGAGCTCCATATGCTGCTGGAGAGCAAGCTACTTTAA 350

QY 121 TTAGCTTCAGAAATGAGATGCCATGAAATTTGCCAGTTGATGATGAGAGCTGCT 180
DB 351 TTGACTTCACGGAGAGGGGGCCATTGAGTTGAGACAGCGAGATCTCCAGTGGACT 410
QY 181 GCTGTGCCCCGAGG 194
DB 411 CAGCTCTCAGAGG 424

RESULT 14

US-10-029-386-11265
Sequence 11265, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: A60MCA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11265
LENGTH: 541
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO 297205.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
OTHER INFORMATION: SWISSPROT HIT: P22793, EVALU8 1.00e-04
US-10-029-386-11265

Query Match 9.8%; Score 69.2; DB 13; Length 541;
Best Local Similarity 54.3%; Pred. No. 4.9e-12;
Matches 140; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 308 CCCCACTGAGATATGAGAGCCCACTCCGATACGAGCCCACTGAGAGATATG 367
DB 14 CTCACCTGAGAGGCTCTCTCTCTCACTGAGGCTCTCTCTCACTGAGGCTCTT 73
QY 368 GAGCCCAACCGTATGAAATGAGAGCCCGCTGTGGATACAGAGCTCACTGTGCAT 427
DB 74 CTCTCCACCTGAGAGGCTCTCTCTCTCACTGAGGCTCTCTCTCACTGAGGCT 133
QY 428 ATGAGGCCCACTCTGATACGAGAGCCCACTGAGATATGAGAGCCCACTCTAG 487
DB 134 CTCTCTTCCACCTGAGAGGCTCTCTCTCTCACTGAGGCTCTCTCTCTCACTGAG 193
QY 488 GATATGAGAGCCCACTCTGATATGAGAGCCCACTGAGATATGAGAGCCCACTCT 547
DB 194 GCTCCCTCTCTCACTGAGAGGCTCTCTCTCTCACTGAGGCTCTCTCTCTCACTG 253
QY 548 TCGATATGAGAGCCCACT 565
DB 254 CAGGCTCTCTGAGAGCC 271

RESULT 15

US-10-029-386-24981/c
Sequence 24981, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: A60MCA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20

```
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO: 24981
; LENGTH: 511
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO 297205.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
US-10-029-386-24981
```

```
Query Match          9.6%; Score 67.8; DB 13; Length 511;
Best Local Similarity 50.8%; Pred. No. 1.4e-11;
Matches 162; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
```

```
QY 313 CTTGCAGATATGAGAGCCCACTCCCGATACGAGCCCACTGCAGATATGAGCC 372
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 343 CTTGCAGGTGAGAGAGAGAGCTTGCAGGTGAGAGAGAGAGAGCTGCAGGTGAGAG 284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 373 CAACCCGTAGAAATGAGCCCGCGCTGTGGATACAGAGCCTCACTGTGGATATGGA 432
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 283 GAGCTGTGAGGTGAGAGAGAGAGAGCTTGCAGGTGAGAGAGAGAGAGCTTGCAGGTGAGAG 224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 433 GCCCACTCTTGGATACGAGAGCCCACTGCAGATATGAGAGCCCACTTAGATAT 492
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 223 GGGAGAGCTTGCAGGTGAGAGAGAGAGAGCTTGCAGGTGAGAGAGAGAGAGCTTGCAGGTGAG 164
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 493 GGAGCCCACTTGTGATATGAGAGCCCACTTGTGATATGAGAGCCCACTTGTGGA 552
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 163 GGAGGGAGAGCTTGCAGGTGAGAGAGAGAGAGAGCTTGCAGGTGAGAGAGAGAGAGAGCTTGCAGGT 104
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 553 TATGAGAGCCCACTTGCAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 612
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 103 GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 44
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 613 GATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 631
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 43 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 25
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: December 16, 2003, 17:58:46
Job time : 233.016 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 09:21:45 ; Search time 1450.89 Seconds
(without alignment)
11809.735 Million cell updates/sec

Title: US-09-864-291-11_COPY_1_705

Perfect score: 705
Sequence: 1 atgcattgctcgtcgtac.....cctcttcagtcctcctc 705

Scoring table: IDENTITY_NUC
Gapex 10.0, Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthm:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gd_estc1:*
10: gd_estc2:*
11: gd_estc3:*
12: gd_estc4:*
13: gd_estc5:*
14: gd_estc6:*
15: em_esthm:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vtc:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gd_gss1:*
29: gd_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	656.4	93.1	1079	12	BMS64167
2	368.2	52.2	663	10	BG699398
3	362.8	51.5	775	10	BG722815
4	333.2	47.3	1414	11	AK015863

Result No.	Score	Query Match	Length	ID	Description
5	263	37.3	701	10	BG701881
6	252.2	35.8	717	14	BY715414
7	169.8	24.1	814	13	BU961805
8	160.4	22.8	488	28	AQ997933
9	158.8	19.7	426	28	AQ997939
10	117.2	16.6	671	29	AG062842
11	104.2	14.8	464	10	BF000754
12	101	14.3	488	10	BF476944
13	100.6	14.3	568	12	BI390687
14	100.6	14.3	636	12	BM490729
15	98	13.9	499	13	BQ356763
16	97.4	13.8	560	13	BU103740
17	97.4	13.8	700	13	BU284977
18	97.4	13.8	723	9	AJ452591
19	97.4	13.8	744	13	BU230211
20	97.4	13.8	932	13	BU107273
21	90.4	12.8	706	13	BU433499
22	89.8	12.7	773	9	AV899050
23	88.8	12.6	755	13	BM089605
24	88.4	12.5	772	9	AJ454950
25	88	12.5	611	9	AL846063
26	87.6	12.4	686	12	BP021598
27	87.2	12.4	594	12	BM190972
28	87.2	12.4	629	10	BF828725
29	87.2	12.4	816	13	BU916890
30	87.2	12.4	931	13	BU914288
31	86.8	12.3	562	12	BM495739
32	86.6	12.3	622	10	BE468886
33	86.6	12.3	899	10	BE470466
34	86.4	12.3	613	9	AL849211
35	86.4	12.3	635	9	AL889954
36	86.4	12.3	647	9	AL849651
37	86.4	12.3	655	9	AL898723
38	86.4	12.3	661	9	AL972453
39	86.4	12.3	663	9	AL881810
40	86.4	12.3	663	9	AL896682
41	86.4	12.3	735	12	BP019748
42	86.4	12.3	823	13	BM087191
43	85.6	12.1	460	10	BE680329
44	85.6	12.1	509	10	BG515416
45	85.6	12.1	607	12	BM190975

ALIGNMENTS

RESULT 1
BMS64167
LOCUS
DEFINITION
AGENCOURT_6560149 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742129
5', mRNA sequence.
BMS64167
VERSION
BMS64167.1 GI:18811738
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
REFERENCE
1 (bases 1 to 1079)
AUTHORS
Mammalia; Eukarya; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Huteleostomi;
NITH-MGC http://mgc.nci.nih.gov/
COMMENT
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
plate: LHAM12759 row: m column: 10
High quality sequence stop: 726.

FEATURES
source

Location/Qualifiers
1.1079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5742129"
/issue_type="medulla"
/lab_host="DH10B"
/clone_lib="NIH_MGC_119"
/note="Organ: Brain; Vector: pCMV-Sport6; Site: 1; Nct1; Site: 2; Score (decoy): RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (BcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH MGC Library."

BASE COUNT 267 a 297 c 268 g 247 t
ORIGIN

Query Match 93.1%; Score 656.4; DB 12; Length 1079;
Best Local Similarity 97.3%; Pred. No. 1.8e-170;
Matches 689; Conservative 0; Mismatches 16; Indels 3; Gaps 2;

1 ATGCCATTGATCTGATGAGAACTCCTGCTTGAACAACAGTATTGCTGAACTTC 60
269 ATGCCATTGATCTGATGAGAACTCCTGCTTGAACAACAGTATTGCTGAACTTC 328
61 ATTAAGGAAGCTATTCAGGAGCTCCATATGATGCTGGGAGAGAGAACTCTTTAA 120
329 ATTAAGGAAGCTATTCAGGAGCTCCATATGATGCTGGGAGAGAGAACTCTTTAA 388
121 TAGTCTTCAGAAATGAGATGATGATGATGATGATGATGATGATGATGATGATG 180
389 TAGTCTTCAGAAATGAGATGATGATGATGATGATGATGATGATGATGATGATG 448
181 GCTGTGCTCCGAGAGATTTCCATTTAGAACTTAAATGATGCTGCTGATGGAATT 240
449 GCTGTGCTCCGAGAGATTTCCATTTAGAACTTAAATGATGCTGCTGATGGAATT 508
241 TAGTAATTAATCTGGGAGAGAAATATGATGATGATGATGATGATGATGATGATG 300
509 TAGTAATTAATCTGGGAGAGAAATATGATGATGATGATGATGATGATGATGATG 568
301 TATGAGGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGCCCACTGCA 360
569 TATGAGGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGCCCACTGCA 628
361 GATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGCCCACTG 420
629 GATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGCCCACTG 688
421 GTCGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAG--CCCC 479
689 GTCGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGCCCC 748
480 ACCTCTAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCTCGATATGAGAG 539
749 ACCTCTAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCTCGATATGAGAG 808
540 CCCACCTCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCTCGAGATATG 599
809 CCCACCTCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCTCGAGATATG 868
600 AGCTCACTCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCTCGAGATATG 657
869 AGCTCACTCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCTCGAGATATG 928
658 GAAAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
929 GAAAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 976

RESULT 2
BG699398 663 bp mRNA linear EST 07-MAY-2001
602679081P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4811804 5',
DEFINITION
LOCUS
ACCESSION
BG699398
VERSION
BG699398.1 GI:13967653
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 663)
NIH-MGC <http://mgi.mci.nih.gov/>.
Natl. Inst. of Health, Mammalian Gene Collection (MGC)
UNPUBLISHED
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
<http://image.llnl.gov>
Plate: LLM10703 row: a column: 21
High quality sequence stop: 660.

FEATURES
sourceLocation/Qualifiers
1.663/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4811804"
/issue_type="hippocampus"
/lab_host="DH10B"/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtagag
); Oligo-dT primed using primer 5'-TTTCTTTTCTTTTCTT-3',
size-selected for average insert size 2.5 kb and
normalized to RQF 5. This is a primary library enriched
for full-length clones and constructed using the
cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."BASE COUNT
ORIGIN

168 a 158 c 169 g 164 t 4 others

Query Match 52.2%; Score 368.2; DB 10; Length 663;
Best Local Similarity 97.1%; Pred. No. 7.6e-91;
Matches 395; Conservative 0; Mismatches 10; Indels 2; Gaps 2;

1 ATGCCATTGATCTGATGAGAACTCCTGCTTGAACAACAGTATTGCTGAACTTC 60
269 ATGCCATTGATCTGATGAGAACTCCTGCTTGAACAACAGTATTGCTGAACTTC 317
258 ATGCCATTGATCTGATGAGAACTCCTGCTTGAACAACAGTATTGCTGAACTTC 317
61 ATTAAGGAAGCTATTCAGGAGCTCCATATGATGCTGGGAGAGAGCAAGCTACTTTAA 120
318 ATTAAGGAAGCTATTCAGGAGCTCCATATGATGCTGGGAGAGAGCAAGCTACTTTAA 377
121 TAGTCTTCAGAAATGAGATGATGATGATGATGATGATGATGATGATGATGATG 180
378 TAGTCTTCAGAAATGAGATGATGATGATGATGATGATGATGATGATGATGATG 437
181 GCTGTGCTCCGAGAGATTTCCATTTAGAACTTAAATGATGCTGCTGATGGAATT 240
438 GCTGTGCTCCGAGAGATTTCCATTTAGAACTTAAATGATGCTGCTGATGGAATT 497
241 TAGTAATTAATCTGGGAGAGAAATATGATGATGATGATGATGATGATGATGATG 300
498 TAGTAATTAATCTGGGAGAGAAATATGATGATGATGATGATGATGATGATGATG 557

Oy	301	TATGGGCGCCACCTGCAGATTATGAGACCCCACCTCCGGATTAGAGAGCCCACTGCAGA	360
Dd	558	TATGAGGCCCACTGCAGATTATGAGACCCCACCTCCGGATTAGAGAGCCCACTGCAGA	615
Oy	361	GGATTATGAGAGCCCAACCGGTAGAAATGAGAGCCCGCCTGTGGATA	407
Dd	616	GGATTATGAGAGCCCAACCGGTAGAAATGAGAGCCCGCCTGTGGATA	662
RESULT 3			
Bg722815		775 bp	mRNA
LOCUS			linear
DEFINITION			EST 08-MAY-2001
Accession	Bg722815	602695278r1 NIH_MGC_97 Homo sapiens CDNA clone IMAGE:4827427 5', mRNA sequence.	
Version	Bg722815		
Keywords	Bg722815.1 GI:14002002		
Source	EST.		
Organism	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
Authors	1 (bases 1 to 775)		
Title	NIH-MGC http://mgc.nci.nih.gov/.		
Journal	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished		
Comment	Contact: Robert Strausberg, Ph.D. Email: cgsaps-remail.nih.gov Tissue Procurement: Miklos Pavlovitz, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN) cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov Plate: LILAM10743 row: 1 column: 20 High quality sequence stop: 651.		
FEATURES			
Source	location/Qualifiers		
	1..775		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4827427"		
	/lab_host="DH10B"		
	/clone_id="NIH_MGC_97"		
	/note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgaggg...); Oligo-dT primed using primer 5'-TTTATTTTTTTTTCVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapped method (Carninci, In preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."		
BASE COUNT	191 a	212 c	193 g
ORIGIN		179 t	
Query Match	51.5%	Score 362.8;	DB 10; Length 775;
Best Local Similarity	95.1%;	Pred. No.2=89;	
Matches	428; Conservative	0; Matches 17;	Indels 5; Gaps 5;
Oy	1	ATGCCATTGATCTGATGACGAACCTCACTGTTGAACAACAGCATTTTGCTGTCAAATCTC	60
Dd	276	ATGCCATTGATCTGATGACGAACCTCACTGTTGAACAACAGCATTTTGCTGTCAAATCTC	335
Oy	61	ATTAAAGGAACCTATTCAAGCAGACTCCATATGCTGCTGGAAAGACAAGCTAATTTTAA	120
Dd	336	ATTAAAGGAACCTATTCAAGCAGACTCCATATGCTGCTGGAAAGACAAGCTAATTTTAA	395
Oy	121	TTAGCTTCAGAAAATGAGATGTCATTGAATTTTGCCAGTTGATGCTGTAAGAGCTGCTCT	180
Dd	396	TTAGCTTCAGAAAATGAGATGTCATTGAATTTTGCCAGTTGATGCTGTAAGAGCTGCTCT	455

QY	181	GCCTTTCGCCCAAGGATTTTCCACTTAGAAGACCTTAA-1GACTGCTTACGCTCTAATGGAAAT	239
Db	456	GTCTCTGCCCGAGATTTTCCACTTAGAAGACCTTAACTBACTGCTTACGCTCTAATGGAAAT	515
QY	240	TTATGTAATTACTGCGGGAAGGGAATATGTGCACTCCAAGATGCCCTTGTTCAGTTATGT	299
Db	516	TTATGTAATTACTGCGGGAAGGGAATATGTGCACTCCAAGATGCCCTTGTTCAGTTATGT	575
QY	300	CTATGGGGCCCCCACTCTGCAAG-ATATGGAAGCCCCCACTTCCCGGATATAGAGACCCCACTG	358
Db	576	CTATGGAAGCCCCCACTCTGCAAGCAATATGAGGCCCACTTCCCGGATATAGAGACCCCACTG	634
QY	359	CAGATATGAGACCCCACTCTGCAAGCAATATGAGGCCCGCTGTGGGATATACAGAGCTC-A	417
Db	635	CAGATATGAGACCCCACTCTGCAAGCAATATGAGGCCCGCTGTGGGATATACAGAGCTC-A	694
QY	418	CTGTGTGCGA-TATGAGAGCCCCCACTCTTGG	446
Db	695	CTGTGTGCGA-CATGAGAGCCCACTCTCGG	724
RESULT 4			
AK015863			
LOCUS			
DEFINITION			
ACCESSION	AK015863	1414 bp	mRNA
VERSION	AK015863.1		
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE			
1	Carninci, P. and Hayashizaki, Y.		
2	High-efficiency full-length cDNA cloning		
3	Meth. Enzymol. 303, 19-44 (1999)		
4	Normalization and subtraction of cap-trapper-selected cDNAs to		
5	prepare full-length cDNA libraries for rapid discovery of new genes		
6	Genome Res. 10 (10), 1617-1630 (2000)		
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			
91			
92			
93			
94			
95			
96			
97			
98			
99			
100			

ORGANISM *Homo sapiens*
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 701)
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.

Tissue Procurement: Mkllos Malkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshituki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Place: LNL0713 row: 1 column: 10
High quality sequence stop: 680.

FEATURES	Location/Qualifiers
source	1. .701

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4815843"
/rnause_type="hippocampus"
/lab_host="DH10B"
/clone_id="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTNN-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

```

Query Match	Similarity	Score	DB	Length
Best Local Similarity	89.4%	Pred. No. 1e-61		
Matches	387	Conservative	0	Mismatches 30; Indels 16; Gaps 9

QY	1	ATGCCATTGATCTGATGACCAACCTC--ACTGTTGAACAACAGATATTGCT---GCAA	55
Db	261	ATGCCATTGATCTGATGACCAACCTCTTAACTGTTGAACAACAGATATTGCTCTCGAAA	320
QY	56	ACTTCATTAAGGGAACATTAATCAGAGAGCTCCAATATGAGCTGCGGGAAGAGACAAAGTACTT	115
Db	321	CTTGCAATTAAGGGAACATTAATCAGAGAGCTCCCAATATGAGCTGCGGGAAGAGACAAAGTACTT	380
QY	116	TT----AAATTAGCTTCAGAAATGAGATGCCATTGAAATTTGCCAGTTGATGTGTGAAA	171
Db	381	TTATTAATAATGATGATCTTCAGAAATGAGAGTGCCATTGAAATTTGCCAGTTGATGTGTGAAA	440
QY	172	GCTGCTCTGCTGTGCTGCCCGAGAGATTTTCCACTTAACCTTAA--ATGACTGGTTCAAGTCT	230
Db	441	GCGGCTCTGCTGCTGCCCGAGAGATTTTCCACTTAAACCTTAAAGATGACTGGTTCAAGTCT	500
QY	231	TATGGAAATTATGTAATTACTTGAGGGAAGGAAATGTGCA--CTCCACAGATGCTTGT	288
Db	501	TATGGGAATTTATGTAATTACTTGAGGGAATGTGCAATCGTCCACAGATGCTTGT	560
QY	289	TCAGTTATTGTCTATGGGCCCCCACTTGAGGATATGAGGCCCCCACTTCCGGATACGGA	348
Db	561	ACAGTTA--TGCTCTATGAGAGCGCACCTGCGGATATGAGGCCCCCACTT--CCGGAATACGGA	618
QY	349	GCCCCACCTGAGAGATATGAGGCCCAACCCGTGAGAAATGAAGGCCGCTTGCGGATAC	408
Db	619	GCCCCA--CTGCAAGATATGAGGCCCAACCCGTGAGAAATGAAGGCCG--CTGTGCGGATAC	676
QY	409	AGAGCTTACCTG	421

Db 677 AGAGCCTCACCTG 689

RESULT 6	BY715414	717 bp	mRNA	linear	EST 17-DEC-2002
LOCUS	BY715414				
DEFINITION	BY715414 RIKEN full-length enriched, adult male testis Mus musculus				
ACCESSION	CCNA Clone 4930521123 5', mRNA sequence.				
VERSION	BY715414				
KEYWORDS	BY715414.1 GI:27128531				
SOURCE	EST.				
ORGANISM	Mus musculus (house mouse)				
	Mus musculus				

REFERENCES

TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	22354683
PUBMED	12466851
COMMENT	Contact: Yoshihide Hayashizaki

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-researc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda
 S., Hashizume, M., Hayashizaki, K., Hirozane, T., Hoti, F., Imocani, K.,
 Ishii, Y., Itoh, M., Kageawa, I., Kawai, J., Kojima, Y., Kondō, S., Komio
 H., Koyas, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,
 Numazaki, R., Ono, M., Osato, N., Shitko, R., Sakazume, N., Sano, H.,
 Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y.,
 Waki, K., Watabiki, A., Muramatsu, M. and Hayashizaki, Y. Direct
 Substitution
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multichannel sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Db	243	CCTCTGGATCAAGGGCCACGTCGTGTAGACACAGGAAGTCCACCTGCTGGGTCTGAAGA	302
Qy	625	AGGCTTCAAGAAATCTACACAGCCGAGGCTCTGTAAGACAGGCTTCTCTTCCCTGCCC	684
Db	303	GGCATTCCCAATCTGTGGCAGTCCAGAACTCTGAATTCAGGATCTTTTCCCTTAC	362
Qy	685	TCCTCTCTCAGGTCCATCTT	705
Db	363	TCGTCTCAAGATCCACTCT	383
RESULT 8			
LOCUS	AQ997933/c	488 bp	DNA linear GSS 24-FEB-2000
DEFINITION	RPCT-23-384A20.TU RPCT-23 Mus musculus genomic clone RPCT-23-384A20		
ACCESSION	RPCT-23-384A20		
KEYWORDS	'genomic survey sequence.'		
SOURCE	AQ997933		
ORGANISM	AQ997933.1 GI:7072953		
REFERENCE	GSS.		
AUTHORS	Mus musculus (house mouse)		
TITLE	Mus musculus		
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
COMMENT	1 (bases 1 to 488)		
	Zhao,S., Nieman,W., Feldblyum,T., Malek,U., Shatman,S., Akinet		
	'B., Levin,M., McGinn,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.		
	Mouse BAC End Sequences from Library RPCT-23		
	Unpublished		
	Other GSSs: RPCT-23-384A20.TU		
	Contact: Shaying Zhao		
	Department of Eukaryotic Genomics		
	The Institute for Genomic Research		
	9712 Medical Center Dr., Rockville, MD 20850, USA		
	Tel: 301 838 0200		
	Fax: 301 838 0208		
	Email: zhaosh@igf.org		
	Clones are derived from the mouse BAC library RPCT-23. For BAC		
	library availability, please contact Pieter de Jong		
	(pieter@djong.med.buffalo.edu). Clones may be purchased from		
	BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)		
	or from Resea ch Genetics (info@resgen.com). BAC end page:		
	http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html		
	Plate: 384 row: A column: 20		
	Seq primer: SP6		
	Class: BAC ends.		
FEATURES			
source	location/Qualifiers		
	1..488		
	/organism="Mus musculus"		
	/mol_type="genomic DNA"		
	/strain="C57BL/6J"		
	/db_xref="taxon:10090"		
	/clone="RPCT-23-384A20"		
	/sex="Female"		
	/lab_host="DH10B"		
	/clone_idb="RPCT-23"		
	/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1:		
	ECORI; Site 2: ECORI; Female C57BL/6J mouse kidney and/or		
	brain genomic DNA was isolated and partially digested		
	with a combination of EORI and EORI Methylase. Site		
	selected DNA was cloned into the pBAC3.6 vector at the		
	ECORI sites. The ligation products were transformed into		
	DH10B electrocompetent cells (BRL life Technologies). "		
BASE COUNT	102 a 128 c 157 g 100 t	1 others	
ORIGIN			
Query Match	22.8%;	Score 160.4;	DB 28; Length 488;
Best Local Similarity	67.3%;	Pred. No. 2e-33;	
Matches 243; Conservative	0;	Mismatches 112;	Indels 6; Gaps 14;
Qy	289	TCAGTATTATCTATAGGGGCCCCACCTGACAGATATGAGACCCCACTCCCGATATACGA	348
Db	457	TACCCCGGTAGATACGAGATCCCATCTGACGGATATGAGAGCCCACTGCGGGATACGA	398

FEATURES	source
LOCUS	AO997939/c
DEFINITION	AO997939 426 bp DNA linear GSS 24-FEB-2000 RPCI-23-384A21.TV RPCI-23 Mus musculus genomic clone RPCI-23-384A21
ACCESSION	AO997939
VERSION	AO997939.1
KEYWORDS	GI:7072959
SOURCE	GSS.
ORGANISM	Mus musculus (house mouse)
REFERENCE	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 426)
AUTHORS	Zhao, S., Nieman, W., Feldblyum, T., Malek, J., Shatsman, S., Aktiret, B., Levins, M., McGam, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Frazer, C.M.
TITLE	Mouse BAC End Sequences from Library RPCI-23
JOURNAL	Unpublished
COMMENT	Other GSSs: RPCI-23-384A21.TV Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (lifesciences.com). BAC end page: http://www.tigr.org/tdb/bac ends/mouse/bac_end_intro.html Plate: 364 row: A column: 21 Seq primer: SP6 Class: BAC ends.
FEATURES	location/Qualifiers
source	1..426 /organism="Mus musculus" /mol_type="genomic DNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="RPCI-23-384A21" /sex="Female" /lab_host="DH10B" /clone_1id="RPCI-23" /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRII. Size selected DNA was cloned into the pGEM3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT

88 a 107 c 134 g 97 t

Query Match 19.7%; Score 138.8; DB 28; Length 426;
Best Local Similarity 66.9%; Pred. No. 1.7e-27;
Matches 230; Conservative 0; Mismatches 107; Indels 7; Gaps 2;

QY 307 GCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 366
DB 425 GTCCATCTGACAGGATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 366
QY 367 GAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 426
DB 365 GTAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 312
QY 427 TATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 486
DB 311 TATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 252
QY 487 GATATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 546
DB 251 GATATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 192
QY 547 CTGAGATATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 605
DB 191 ATGAGATATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 132
QY 606 ACTGCTGATATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 649
DB 131 ACTGCTGATATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 88

RESULT 10

AG062842 671 bp DNA linear GSS 03-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-051F01.R, genomic survey sequence.
AG062842
VERSION AG062842.1 GI:16614644

KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes

REFERENCE 1
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totsuki, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished

AUTHORS 2 (bases 1 to 671)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totsuki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimbas@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)

COMMENT Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS
Sequencing: MJ3Rev
LIBRARY Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers 1. .671

FEATURES
source

/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-051F01.R"
/sex="male"
/cell_type="lymphoblast"
/clone_11b="PTB Chimpanzee Male BAC library"

BASE COUNT

77 a 270 c 103 g 220 t 1 others

Query Match 16.6%; Score 117.2; DB 29; Length 671;
Best Local Similarity 55.6%; Pred. No. 2.1e-21;
Matches 224; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 297 TGTCTATGAGGCCCCCACTGACGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 356
DB 196 TGTCTATGAGGCCCCCACTGACGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 255
QY 357 TGTCTATGAGGCCCCCACTGACGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 416
DB 256 TGTCTATGAGGCCCCCACTGACGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 315
QY 417 ACCTGTGATATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 476
DB 316 TGTCTATGAGGCCCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 375
QY 477 CCCACTCTGATATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 536
DB 376 TGTCTATGAGGCCCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 435
QY 537 AGCCCACTCTGATATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 596
DB 436 AGCCCACTCTGATATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 495
QY 597 CAGAGCTCTGATATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 656
DB 496 CAGAGCTCTGATATGAGCCCACTGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 555
QY 657 TGTCTATGAGGCCCCCACTGACGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 699
DB 556 TGTCTATGAGGCCCCCACTGACGATATGAGCCCACTCCGATGACGAGCCCACTGAGGATAT 598

RESULT 11

BP000754 464 bp mRNA linear EST 06-OCT-2000
LOCUS BP000754
DEFINITION 7143901.x1 NCI CGAP B16 Homo sapiens cDNA clone IMAGE:3337488 3'

KEYWORDS similar to contains element TARI repetitive element ;, mRNA
SOURCE Homo sapiens (human)
EST. BP000754.1 GI:10701029

ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 464)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Dennis Sgroi, M.D., Kristina Cole, M.D., Ph.D.
Student, Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
found through the NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 358.

FEATURES

source

Location/Qualifiers

1..464
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3337488"
/sex="female"
/tissue_type="lobular carcinoma in situ"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI CGAP Br16"
/note="Organ: breast; Vector: pAMP1; mRNA made from breast carcinoma tissue, cDNA made by oligo-dT priming. Directionally cloned. Size-selected on agarose gel, average insert size 400 bp. Primary library, non-amplified."

BASE COUNT 58 a 202 c 59 g 145 t
ORIGIN

Query Match 14.8%; Score 104.2; DB 10; Length 464;
Best Local Similarity 53.9%; Pred. No. 6.8e-18;
Matches 214; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

QY 306 GAGCCCACTGAGATATGAGAGCCCACTCCGATACGAGCCCACTGAGATAT 365
DB 58 GGCAGCTTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTCA 117
QY 366 TGGAGCCCACTGAGATATGAGAGCCCGCTGTGAGATACAGAGCTTCACTGTGCG 425
DB 118 CACAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTT 177
QY 426 ATATGAGAGCCCACTGTTGATATGAGAGCCCACTGAGATATGAGAGCCCACTCT 485
DB 178 TCAGATAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTCC 237
QY 486 AGATATGAGAGCCCACTCTGTTGATATGAGAGCCCACTCTGTTGATATGAGAGCCCACT 545
DB 238 TGTTCACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCC 297
QY 546 TCTCGATATGAGAGCCCACTCTGTTGATATGAGAGCCCGCTGTGAGATACAGAGCTCTC 605
DB 298 TCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTC 357
QY 606 ACCGCTGATATGAGAGCCCACTCTGTTGATATGAGAGCCCACTCTCTGTTGATATGAGAG 665
DB 358 TCTCATATGATATGAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCT 417
QY 666 GAGTCTCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTCA 702
DB 418 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 454

RESULT 12
BF476944/c 488 bp mRNA linear EST 05-DEC-2000
LOCUS na556d10.x1 NCI CGAP Br16 Homo sapiens cDNA clone IMAGE:3260227 3'
DEFINITION similar to contains element YARI repetitive element ;, mRNA
sequence.

ACCESSION BF476944
VERSION BF476944.1 GI:11547771
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bld-research.nih.gov
Tissue Procurement: Dennis Sgroi, M.D., Kristina Cole, M.D., Ph.D.

FEATURES

source

Location/Qualifiers

1..488
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3260227"
/sex="female"
/tissue_type="lobular carcinoma in situ"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI CGAP Br16"
/note="Organ: breast; Vector: pAMP1; mRNA made from breast carcinoma tissue, cDNA made by oligo-dT priming. Directionally cloned. Size-selected on agarose gel, average insert size 400 bp. Primary library, non-amplified."

BASE COUNT 154 a 68 c 204 g 62 t
ORIGIN

Query Match 14.3%; Score 101; DB 10; Length 488;
Best Local Similarity 53.1%; Pred. No. 5.4e-17;
Matches 215; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 297 TGTCTATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATACGAGCCCACTGAGATAT 356
DB 479 TGTTCACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCC 420
QY 357 TGGAGATATGAGAGCCCACTGTTGATATGAGAGCCCGCTGTGAGATACAGAGCTCTC 416
DB 419 TCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCC 360
QY 417 ACTGTCGATATGAGAGCCCACTCTGTTGATATGAGAGCCCACTCTGTTGATATGAGAGCT 476
DB 359 TCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCC 300
QY 477 CCACTCTGATATGAGAGCCCACTCTGTTGATATGAGAGCCCACTCTCTGTTGATATGAGAG 536
DB 299 CTCCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCT 240
QY 537 AGCCCACTCTGATATGAGAGCCCACTCTGTTGATATGAGAGCCCGCTGTGAGATACAGAG 596
DB 239 AGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
QY 597 CAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 656
DB 179 CACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
QY 657 TGAAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 701
DB 119 TCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 75

RESULT 13
B1390687 568 bp mRNA linear EST 06-AUG-2001
LOCUS pppln.pk001.p22 Normalized chicken pituitary/Hypothalamus/plinal
DEFINITION library Gallus gallus cDNA clone pppln.pk001.p22 5' similar to
gb|U09511.1 (U09458) WW domain binding protein-2 [Homo sapiens].
mRNA sequence.

ACCESSION B1390687
VERSION B1390687.1 GI:15083969
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 568)
Porter,T.B. and Cogburn,L.A.
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA library USDA/IRAFs Animal Genome Project

JOURNAL

Unpublished
Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburnudel.edu, www.chickest.udel.edu.

COMMENT

FEATURES

source

1. 568
Location/Qualifiers

/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Commercial broiler chicken"
/db_xref="taxon:9031"
/clone="pgp2n.pk001.p22"
/sex="Male and Female"
/tissue_type="Pituitary gland/Hypothalamus/Pineal gland"
/dev_stage="Embryonic (dl2,dl4,dl9) ; post-hatch (w1,w3,w5,w7,w9)"
/lab_host="R. coli EMDH10B"
/clone_lib="Normalized Chicken Pituitary/Hypothalamus/Pineal Library"
/note="Vector: PCWVSPORT6; Library made from equivalent pools of total RNA isolated from each tissue at different ages. Single pass sequencing from 5'-end"

BASE COUNT

128 a 146 c 159 g 130 t 5 others

ORIGIN

Query Match 14.3%; Score 100.6; DB 12; Length 568;
Best Local Similarity 68.5%; Pred. No. 7.6e-17;
Matches 139; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

1 ATGCCATTGATCTGATGACGAACTCTGTTGAAACCAAGTATTCCTGCAACTTC 60

236 ATGCCATTGATCTGATGACGAACTCTGTTGAAACCAAGTATTCCTGCAACTTC 295

61 ATTAGGAGAACTATTCAGGAGAGCTCATATGCTGCTGGAGAGCAAGCTACTTTAA 120

296 ATCAAGAGACAGATTGAGCTGAGAGAGAGTGGCTGGAGAGGAGGAAAGCTTTAA 355

121 TTAGCTTCAGAAATGAGATGCGATTGAATTTGCCAGATTGATGCTGAAAGCTGCT 180

356 CTGACTTTCAACAGGAGGAGGAGCATGAGTTTGACAGCTGATGTTCAAGCTGCTCT 415

181 GCTGTTGCCAGAGATTTCCTACT 203

416 AGTGCTTCAGTGAGAGTTCCTCT 438

RESULT 14

BM490729

LOCUS

DEFINITION

BM490729 636 bp mRNA linear EST 07-FEB-2002
pgp2n.pk004.e6 Normalized Chicken Pituitary/Hypothalamus/Pineal library (pgp2n) Gallus gallus cDNA clone pgp2n.pk004.e6 5' similar to gblAD10951.1 (U79458) WW domain binding protein-2 [Homo sapiens], mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Gallus gallus (chicken)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 636)
Porter,T.B. and Cogburn,L.A.
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA library

JOURNAL

Unpublished
Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburnudel.edu, www.chickest.udel.edu.

COMMENT

FEATURES

source

1. 636
Location/Qualifiers

/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Commercial broiler chickens"
/db_xref="taxon:9031"
/clone="pgp2n.pk004.e6"
/sex="Male and Female"
/tissue_type="Pituitary gland/Hypothalamus/Pineal gland"
/dev_stage="Embryonic (dl2,dl4,dl9) ; post-hatch (1,3,5,7,9 weeks)"
/lab_host="R. coli EMDH10B"
/clone_lib="Normalized Chicken Pituitary/Hypothalamus/Pineal Library (pgp2n)"
/note="Vector: PCWVSPORT6; Library made from equivalent pools of total RNA isolated from each tissue at different ages. Single pass sequencing from 5'-end"

BASE COUNT

143 a 174 c 173 g 146 t

ORIGIN

Query Match 14.3%; Score 100.6; DB 12; Length 636;
Best Local Similarity 68.5%; Pred. No. 8.2e-17;
Matches 139; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

1 ATGCCATTGATCTGATGACGAACTCTGTTGAAACCAAGTATTCCTGCAACTTC 60

236 ATGCCATTGATCTGATGACGAACTCTGTTGAAACCAAGTATTCCTGCAACTTC 295

61 ATTAGGAGAACTATTCAGGAGAGCTCATATGCTGCTGGAGAGCAAGCTACTTTAA 120

296 ATCAAGAGACAGATTGAGCTGAGAGAGAGTGGCTGGAGAGGAGGAAAGCTTTAA 355

121 TTAGCTTCAGAAATGAGATGCGATTGAATTTGCCAGATTGATGCTGAAAGCTGCT 180

356 CTGACTTTCAACAGGAGGAGGAGCATGAGTTTGACAGCTGATGTTCAAGCTGCTCT 415

181 GCTGTTGCCAGAGATTTCCTACT 203

416 AGTGCTTCAGTGAGAGTTCCTCT 438

RESULT 15

BQ356763

LOCUS

DEFINITION

BQ356763 499 bp mRNA linear EST 20-MAY-2002
IL3-HT0618-120500-138-A10 HT0618 Homo sapiens cDNA, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 499)
Dias Neto,B., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.P., Goldman,G.H., Carvalho,A.F., Matukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE PUBMED
20202663 10737800
Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the PAPSP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL:
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&cl=IL3-HT0618-120500-138-A10&3=2000-05-12&t=1>)

Seq primer: puc 18 forward

High quality sequence start: 15

High quality sequence stop: 162.

Location/Qualifiers

1..499

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="HT0618"

/note="Organ: head, neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORSTS PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cdna amplification were performed under
low stringency conditions."

BASE COUNT 62 a 216 c 66 g 155 t

ORIGIN

Query Match

Best Local Similarity 13.9%; Score 98; DB 13; Length 499;

Matches 217; Conservative 0; Mismatches 180; Indels 1; Gaps 1;

```
QY 306 GGGCCCACTGACGATATGAGAGCCCACTCCGATACGAGAGCCCACTGACGATA 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44 GGAAGCTTCTCTGTTCAACAGGCTCTCTCTGTTCAAGTAGGCTCTCTCTGTTCA 103
QY 366 TGAAGCCCAACCGTAGAAATGAGAGCCGCTGTGGATACAGAGCTTCACT-GTGC 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 CACAGCTCTCTCTGTTCAAGGCTCTCTCTCTGTTCAACAGGCTCTCTCTCTCT 163
QY 425 GATATGAGAGCCCACTCTTGAATAGGAGCCCACTGACGATATGAGAGCCCACTC 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 TATACATAGCTCTCTCTGTTCAACAGGCTCTCTCTCTGTTCAAGTAGGCTCTCT 223
QY 485 TAGGATATGAGAGCCCACTCTTGAATAGGAGCCCACTCTGATATGAGAGCCCACT 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 CTGTTCAACAGGCTCTCTCTGTTCAAGGCTCTCTCTCTGTTCAAGTAGGCTCTCT 283
QY 545 CTCTCGAATATGAGAGCCCACTGACGAAATGAGAGCCGCTGCGGATACAGAGCT 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 284 CTCTGTTCAAGGCTCTCTCTGTTCAACAGGCTCTCTCTCTGTTCAACAGGCTCT 343
QY 605 CACCTGTGATACAGAGCCGAGGCTCAGAAATACAGAGCCGAGGCTCTGAAACG 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 CTCTCAATGTTCAAGGCTCTCTCTGTTCAACAGGCTCTCTCTCTGTTCAACAGG 403
QY 665 AGGCTTCTCTCTCTGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 441
```

Search completed: December 16, 2003, 17:31:45
Job time : 1452.89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 15:10:30 ; Search time 2514.24 Seconds

(without alignments)
3823.721 Million cell updates/sec

Title: US-09-864-291-12
Sequence: 1 MPPDMLNTLTGQPVFANF.....AQAPENRSLSPASSSQVHS 235

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool.p/US09864291/runat.15122003.160858.23305/app.query.fasta.1.846
-DB=genmb1 -GEMT=fastap -SUFFIX=xgse -MINMATCH=0.1 -LOOPTCL=0 -LOOPEXT=0
-UNITs=bits -START=1 -END=1 -MATRIX=biocum62 -TRANS=humand4 cdi -LIST=45
-DOCALLIGN=200 -THR_SCORES=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODS=LOCAL
-OUTMT=ptc -NORM=ext -HSAFSLZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09864291 @CGN 1 1 3854 @runat.15122003.160858.23305 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEDEV -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Genmb1:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_ey:*
13: gb_un:*
14: gb_vt:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*

29: em_vt:*
30: em_hcg_hum:*
31: em_hcg_in:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vtc:*
38: em_ey:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1279	100.0	1001	6 AX359665	AX359665 Sequence
2	1263	98.7	2267	9 BC022546	BC022546 Homo sapi
3	1235	96.6	2266	9 BC022549	BC022549 Homo sapi
4	777	60.8	1413	6 AX359658	AX359658 Sequence
5	765	59.8	220895	9 HS250D10	Z99716 Human DNA s
6	429	33.5	168425	2 AC113593	AC113593 Mus muscu
7	429	33.5	224086	10 AC104325	AC104325 Mus muscu
8	424	33.2	179222	2 BX296515	BX296515 Sus scrofa
9	424	33.2	222469	2 BX470149	BX470149 Dario rer
10	405	31.7	253149	2 AC107527	AC107527 Rattus no
11	405	31.7	270171	2 AC132969	AC132969 Rattus no
12	383	29.9	129624	2 BX296540	BX296540 Sus scrofa
13	289	22.6	170781	2 AC092212	AC092212 Trypanoso
14	276.5	21.6	2170	9 AB072784	AB072784 Macaca fa
15	272.5	21.3	166045	2 AC134957	AC134957 Tetradon
16	270	21.1	793	10 AF459026	AF459026 Rattus no
17	261.5	20.4	1802	10 MM040826	U40826 Mus muscicu
18	261.5	20.4	1885	6 AR220846	AR220846 Sequence
19	258.5	20.2	198276	2 AC145040	AC145040 Gorilla g
20	258	20.2	196322	2 AC145025	AC145025 Gorilla g
21	257	20.1	1691	9 HS079458	UT9458 Human WW do
22	257	20.1	1820	9 BC007452	BC007452 Homo sapi
23	257	20.1	1835	9 AK057881	AK057881 Homo sapi
24	257	20.1	1912	9 BC010616	BC010616 Homo sapi
25	256	20.0	148851	9 HS155D22	Z97205 Human DNA s
26	255.5	20.0	198833	2 AC144590	AC144590 Gorilla g
27	254.5	19.9	207682	2 AC144462	AC144462 Gorilla g
28	253	19.8	1747	3 AY060828	AY060828 Drosophi
29	252.5	19.7	2557	8 AF255446	AF255446 Cryptheco
30	251	19.6	174921	2 AC012266	AC012266 Homo sapi
31	250	19.5	167067	10 AC124489	AC124489 Mus muscu
32	249.5	19.5	39782	2 AC138635	AC138635 Homo sapi
33	249.5	19.5	40476	2 AC140538	AC140538 Homo sapi
34	244.5	19.1	159064	9 AC138512	AC138512 Homo sapi
35	244	19.1	192533	9 AC055876	AC055876 Homo sapi
36	242.5	19.0	2557	8 AF255446	AF255446 Cryptheco
37	240.5	18.8	198438	2 AC144362	AC144362 Gorilla g
38	239	18.7	151800	2 AC138649	AC138649 Homo sapi
39	236.5	18.5	160174	2 AC091951	AC091951 Homo sapi
40	236.5	18.5	182255	2 AC140810	AC140810 Homo sapi
41	235.5	18.4	199175	2 AC140836	AC140836 Homo sapi
42	234.5	18.3	1205	5 BC052332	BC052332 Dario rer
43	233.5	18.3	141447	9 AC016204	AC016204 Homo sapi
44	233.5	18.3	180233	9 AC011767	AC011767 Homo sapi
45	232	18.1	235381	2 AC103031	AC103031 Rattus no

ALIGNMENTS

RESULT 1

AX359665
LOCUS AX359665 1001 bp DNA linear PAT 13-FEB-2002
DEFINITION Sequence 11 from Patent WO0190185.
ACCESSION AX359665
VERSION AX359665.1 GI:18675411
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Oiko, R. and Sutovsky, P.
TITLE Pt32 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses
thereof
JOURNAL Patent: WO 0190185-A 11 29-NOV-2001;
QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES
UNIVERSITY (US)
FEATURES
source Location/Qualifiers
1..1001
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1..708
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD23337.1"
/db_xref="GI:18675412"
/translation="MPDMLTNLTVBQPVPAANFKIGTIOAAPYGWGCQATFKLVER
NGDAIEFAQIMVLAASAVARGPFLRTIMFSSMGIYVTVGSMNCTPMPSCVIVYG
APPAGGAPPGGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP
PIGAGAPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP
PENASLPSASSQVHS"
BASE COUNT 261 a 254 c 248 g 238 t
ORIGIN
Alignment Scores:
Pred. No.: 1 75e-50 Length: 1001
Score: 1279.00 Matches: 235
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-09-864-291-12 (1-235) x AX359665 (1-1001)
QY 1 MetProPheAspLeuMetThrAsnLeuThrValGluGlnProValPheAlaAsnPro 20
DB 1 ATGCCATTGATGTGATGAGAACCTCAGCTTGAAACAACAGTATTTCCTGCAACTTC 60
QY 21 IIEYEGLYTHRTLEGNAIAAIPRTYRGYLYTTPGLUGLYGNAIAETHRPHELY 40
DB 61 ATTAAGGAACTATTGAGGACGCTCATATGTGCTGGAGAGACAACTACTTTTAA 120
QY 41 IeuaValPheArgAsnGlyAspAlaIleGluPheAlaGlnLeuMetValIysAlaAser 60
DB 121 TTAGTCTTCAGAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 61 AlalaValaArgGlyPheProLeuArgThrLeuAsnAspTTPPheSerSerMetGlyIle 80
DB 181 GCGTGGCCCGAGGATTTCCACTTAGACCTTAAAGACGCTTCAAGCTTATGGGAATT 240
QY 81 TTTTValIleThrGlyGlyGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
DB 241 TATGTAATTAATCGGGAGAGGAAATATGTGACCTCCACATGCTTGTTCAGTTATGTTC 300
QY 101 TTTGValAlaProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAlaProProAla 120
DB 301 TATGGGGCCCGACCTCGAGATATGAGCCCACTCCGAGATACGAGCCCACTCGA 360
QY 121 GATYTYGTYAlaGlnProValGlyAsnGlyGlyProProValGlyTyrArgAlaSerPro 140
DB 361 GATATAGAGCCCAACCTAGAGAAATGAGAGCCGCCCTGTGGGATACAGAGCTCACT 420

QY 141 ValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaPro 160
DB 421 GTGCAATATGAGAGCCCACTCTTGATATGAGAGCCCACTTGACAGATATGAGAGCCCA 480
QY 161 ProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAla 180
DB 481 CCTTAGATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTTGATATGAGAGCC 540
QY 181 ProProLeuGlyTyrGlyAlaProProAlaGlyAsnGlyGlyProProAlaGlyTyrArg 200
DB 541 CCACCTCTCGATATGAGAGCCCACTCGACAGAAATGAGAGCCGCCCTCGCGATACGA 600
QY 201 AlAserProAlaGlySerGlyValaArgProGlnGlnSerThrAlaAlaGlnAlaProGlu 220
DB 601 GCTTACCTGCTGATACAGAGCCCACTCGATATGAGAGCCCACTCTTGATATGAGAGCC 660
QY 221 AsnGluAlaSerLeuProSerAlaSerSerSerGlnValAlaSer 235
DB 661 AACGAGGCTTCTCTTCCCTGCTGCTCTTCTCTGATGCTCATTTCT 705
RESULT 2
BC022546 2267 bp mRNA linear PRI 04-FEB-2002
LOCUS BC022546
DEFINITION Homo sapiens, similar to RIKEN cDNA 4930521I23 gene, clone
MGC:26816 IMAGE:4811804, mRNA, complete cds.
ACCESSION BC022546
VERSION BC022546.1 GI:18490706
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2267)
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NHL-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgapb-re@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLT)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.sbcg.stanford.edu>
Contact: (Dickson, Mark) mcdepaxil@stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNLT at: <http://image.llnl.gov>
Series: IRAX Plate: 32 Row: j Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.
FEATURES
source Location/Qualifiers
1..2267
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26816 IMAGE:4811804"
/tissue_type="Brain, hippocampus"
/clone_id="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
32..961
/codon_start=1
/product="Similar to RIKEN cDNA 4930521I23 gene"
/protein_id="AAH2546.1"

DB	BASE COUNT	709 a	477 c	469 g	612 t
ORIGIN					
Alignment Scores:					
Pred. NO.:	1,81e-49	Length:	2267		
Score:	1263.00	Matches:	232		
Percent Similarity:	98.72%	Conservative:	0		
Best Local Similarity:	98.72%	Mismatches:	3		
Query Match:	98.75%	Indels:	0		
DB:	9	Gaps:	0		
US-09-864-291-12 (1-235) x BC022546 (1-2267)					
QY	1	MecProPheapLeuMetThrAsnLeuThrValGluGlnProValPheAlaAlaAsnDhe	20		
Db	254	ATGCCATTGATCTGATGATGACCACTCACTGTTGAACAACAGATATTGCTGCAAACTTC	313		
QY	21	ILelysgLYTHrILeGlnAlaAsProLYrGlyVGLYrPGluGluGlnAlaThPhelys	40		
Db	314	ATTAAAGGAACTATTCAAGCAGCTCCAAATGCTGCTGGAGAGCAAGCTTAACTTTAA	373		
QY	41	LeuValPheArGAsnGlyAspAlaILeGluPheAlaGlnLeuMetValIysAlaAser	60		
Db	374	TTAGTCTTCAGAAATGAGAGTGCCTTAATTTGCCAGTTGATGATGATAAGCTGCTCT	433		
QY	61	AlaValAlaArGlyPheProLeuArGThrLeuAsnAePTPrPheSerSerMetGlyLe	80		
Db	434	GCTGCTGCCGAGAGATTTCCAATTGAACCTTAATATGACTGGTTCAAGCTTATGGAGAT	493		
QY	81	TyrValILleThrGlyGluGlyAsnMetCyThrProGlnMetProCySerValILleVal	100		
Db	494	TATGTAATTAATCTGGGGAAGGGAATATGCACTCCACAGATGCCCTTTCAGTTAATGTC	553		
QY	101	TyrGlyAlaProProAlaGlyTYrGlyAlaProProProGlyTYrGlyAlaProProAla	120		
Db	554	TATGAGGCCCACTCGCAGAGATATGAGGCCCACTCCCGATATGAGGCCCACTGCA	613		
QY	121	GlyTYrGlyAlaGlnProValGlyVAsnGluGlyProProValGlyTYrArGAlaAserPro	140		
Db	614	GGAATATGAGGCCCAACCCCTGAGAAATGAAGCCCGCTGTGGATATACAGAGCTCACT	673		
QY	141	ValArGlyTYrGlyAlaProProLeuGlyTYrGlyAlaProProAlaGlyTYrGlyAlaPro	160		
Db	674	GTGCGATATGAGGCCCACTCTTGATATACGAGGCCCACTGCGAGATATGAGGCCCA	733		
QY	161	ProLeuGlyTYrGlyAlaProProLeuGlyTYrGlyThrProProLeuGlyTYrGlyAla	180		
Db	734	CCTTATGAGATATGAGGCCCACTCTTGATATGAGAACCCCACTCTGAGATATGAGACC	793		
QY	181	ProProLeuGlyTYrGlyAlaProProAlaGlyVAsnGluGlyProProAlaGlyTYrArG	200		
Db	794	CCACTCTCGGATATGAGGCCCACTGCAAGAAATGAAGCCCGCTGCGGATATGAGA	853		
QY	201	AlaSerProAlaGlySerGlyAlaArGProGlnGlnSerThrAlaAlaGlnAlaProGlu	220		
Db	854	GCTTCACTGCTGATATGAGAGCCAGGCTTCAAGATATACAGAGCCCAAGCTCTGAA	913		
QY	221	AsnGluAlaSerLeuProSerAlaSerSerSerGlnValIAser	235		
Db	914	AACGAGGCTCTCTTCCCTGCTCTCTTCTCAAGTCCAACTTC	958		
RESULT 3					
LOCUS	BC022549	2266 bp	mRNA	linear	FRI 04-FEB-2002
DEFINITION	Homio bap1ens, similar to RIKEN cDNA 4930521.23 gene, clone				
	MG:26828 IMAGE:4815849, mRNA, complete cds.				

ACCESSION	BC022549
VERSION	BC022549.1 GI:18490710
KEYWORDS	MGC.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 2266) Strauberg,R. Direct Submission Submitted (01-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: gcgaps-remail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toehyuki and Piero Carninci (RIKEN) CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www-sbg.stanford.edu Contact: (Dickson, Mark) mcd@pxl.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Series: IRAX plate; 32 Row; n Column: 3 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein. Location/Qualifiers 1..2266 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:26828 IMAGE:4815849" /tissue_type="Brain, hippocampus" /clone_lib="NIH_MGC_95" /lab_host="DH10B" /note="Vector: pBluescript" 32..634 /codon_start=1 /product="Similar to RIKEN cDNA 4930521I23 gene" /protein_id="AAH2549.1" /db_xref="GI:18490711" /translation="MAVNSHTSNRRGALIPNGESILKRSPNVELSPFORSBSNVES GRRTGLFTLSYRVIFPITSCSIDPMISFPMPDPLMTNLTVEQPVPAANI KGTIOAA PYGMEGQATFKLVFRNGAIEFAOIAWKAASAAGAPLPRLTNDWSSNGIVYTISK GICAHRCVLQLSMRPHLODMRPHLPTEPHILODMRNP"
CDS	
BASIS COUNT	709 a 477 c 468 g 612 t
ORIGIN	
Alignment Scores:	
Pred. No.:	3.38e-48 Length: 2266
Scores:	1235.00 Matches: 231
Percent Similarity:	98.30% Conservative: 0
Best Local Similarity:	98.30% Mismatch: 4
Query Match:	96.56% Indels: 1
DB:	Gaps: 9
US-09-864-291-12 (1-235) X BC022549 (1-2266)	
Oy	1 MetProheapleuewcttThrahtLeuthVaJcluglnProvaIpheaJaalaenPh 20
Dd	254 ATGCCATTTCATCTGAAGAACAACCTCACTGTGAAACAACAGATTTGCGCAACTTC 313
Oy	21 IlleyagelyhrrllleglnalaaaprotyglcgltTtpdluglygdlnaatrPhetys 40
Dd	314 ATTAAAGGAACATTATCCAGCAGCTCCATAATAGTGATGGCGAAGACAAGCTACTTTAAA 373

gene similar to neuronal-specific septin 3, a pseudogene similar to AN2 (adenine nucleotide translocator 2), 2 mRNAs based on ESTs, a genomic marker D22S1178, a CA repeat polymorphism, ESTs and a CpG island, complete sequence.

ACCESSION 299716 GI:4456457
 VERSION 299716.4
 KEYWORDS HTG; AN2; CpG island; D22S1178; NAGA; septin 3; SRBP2.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 220895)

REFERENCE 1 (bases 1 to 220895)
 AUTHORS Clark, G.
 TITLE Direct Submission
 JOURNAL Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

COMMENT On Mar 21, 1999 this sequence version replaced gi:4164339. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMBASE; Information on the WORMBASE database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormbase CTA-250D10 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218.

VECTOR: pBAC108L

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk

 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed from the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

This sequence is the entire insert of clone CTA-250D10. The true left end of clone RP1-18601 is at 129979 in this sequence. The true right end of clone RP3-359J16 is at 1335 in this sequence. The true right end of clone RP5-821D11 is at 23458 in this sequence. The true right end of clone RP3-359J16 is at 118711 in this sequence.

FEATURES
 Source

1..220895
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="22"
 /clone="CTA-250D10"
 /clone_1lb="CIT978SK-A1"
 17..29
 /note="2.6 copies 5 mer ATTGG 26% conserved"
 repeat_region
 33..325
 /note="AluX repeat: matches 1..301 of consensus"
 repeat_region
 326..338
 /note="2.6 copies 5 mer TTGGA 26% conserved"
 repeat_region
 667..683
 /note="2.1 copies 8 mer AGAGACC 34% conserved"
 repeat_region
 740..751
 /note="2.4 copies 5 mer TTTTA 24% conserved"

repeat_region 1443..1457
 /note="7.5 copies 2 mer AC 21% conserved"
 repeat_region 1715..1997
 /note="AluDo repeat: matches 5..293 of consensus"
 repeat_region 2103..2378
 /note="AluSg repeat: matches 2..296 of consensus"
 repeat_region 2844..2855
 /note="12.0 copies 1 mer A 24% conserved"
 repeat_region complement(2884..3033)
 /note="MIR repeat: matches 70..235 of consensus"
 repeat_region complement(3040..3344)
 /note="AluY repeat: matches 1..310 of consensus"
 repeat_region 3446..3681
 /note="L1MB4a repeat: matches 5256..5471 of consensus"
 repeat_region 3746..3837
 /note="L1MB4a repeat: matches 5615..5707 of consensus"
 repeat_region 3838..4142
 /note="AluDo repeat: matches 2..305 of consensus"
 repeat_region 4143..4441
 /note="L1MB4a repeat: matches 5707..5939 of consensus"
 repeat_region 4442..4747
 /note="AluSg repeat: matches 1..292 of consensus"
 repeat_region 4748..4858
 /note="L1MB4a repeat: matches 5939..6110 of consensus"
 repeat_region 4881..5172
 /note="AluX repeat: matches 1..296 of consensus"
 repeat_region 5418..5429
 /note="2.4 copies 5 mer GCTGA 24% conserved"
 repeat_region complement(5769..6021)
 /note="match: G8S: Em:A0069839"
 repeat_region 6074..6089
 /note="2.0 copies 8 mer CCGGCGTG 32% conserved"
 repeat_region 6139..6153
 /note="2.5 copies 6 mer TTTAA 21% conserved"
 repeat_region 6141..6155
 /note="3.0 copies 5 mer TTTAA 21% conserved"
 repeat_region complement(6366..6572)
 /note="MIR repeat: matches 5..214 of consensus"
 repeat_region 6839..6852
 /note="2.3 copies 6 mer GGAAT 28% conserved"
 repeat_region 7088..7114
 /note="6.8 copies 4 mer TTTA 36% conserved"
 repeat_region 7091..7109
 /note="1.9 copies 10 mer ATTATTTAC 38% conserved"
 repeat_region complement(7122..7418)
 /note="AluX repeat: matches 1..300 of consensus"
 repeat_region 7464..7502
 /note="19.5 copies 2 mer AT 78% conserved"
 repeat_region 7502..7524
 /note="11.5 copies 2 mer AC 46% conserved"
 repeat_region 7523..7538
 /note="2.0 copies 8 mer CAAATAT 32% conserved"
 repeat_region 7535..7563
 /note="14.5 copies 2 mer AT 49% conserved"
 repeat_region complement(7564..7859)
 /note="AluY repeat: matches 1..296 of consensus"
 repeat_region 7873..7893
 /note="5.2 copies 4 mer TGT 33% conserved"
 repeat_region 7914..8200
 /note="AluDb repeat: matches 1..290 of consensus"
 repeat_region complement(8202..8318)
 /note="AluDo/FLM repeat: matches 1..117 of consensus"
 repeat_region 8387..8409
 /note="11.5 copies 2 mer AC 46% conserved"
 repeat_region 8483..8498
 /note="2.0 copies 8 mer GTGTGTGT 32% conserved"
 repeat_region 8486..8496
 /note="2.2 copies 5 mer GTGCT 22% conserved"
 repeat_region 8606..8622
 /note="2.4 copies 7 mer AGAGAA 34% conserved"
 repeat_region complement(8767..8989)
 /note="MER115 repeat: matches 33..268 of consensus"
 repeat_region 9004..9014

```

"note="2.2 copies 5 mer CTTGG 22% conserved"
9037. 9140
/note="MIR repeat: matches 17. .123 of consensus"
9509. .9823
/note="A1usX repeat: matches 1. .312 of consensus"
9872. .9883
/note="2.4 copies 5 mer GTCCT 24% conserved"
10071. 9606
/gene="SREBF2"
join1<10071.10520,11851.12032,14129.14275,17038.17259
118577.18691,18783.18964,20469.20661,21182.21363,
23956.24232,28082.28251,36357.36525,38060.38177,
40292.40401,41889.42021,43570.43738,46210.46395,
48103.48214,48680.49606)
/gene="SREBF2"
/product="bK25OD10.1 (sterol regulatory element binding
transcription factor 2)"
/note="sterol regulatory element binding transcription
factor 2
match: cDNAB: Em:U22031 Em:U2330 Em:U0966 Em:U09103
Em:U16995 Em:U22818 Em:U22819 Em:U2329 Em:AB017337
match: EST: Em:R33312 Em:AL040138 Em:H92808 Em:AA053866"
/evidence=not experimental
join1<10071.10520,11851.12032,14129.14275,17038.17259
118577.18691,18783.18964,20469.20661,21182.21363,
23956.24232,28082.28251,36357.36525,38060.38177,
40292.40401,41889.42021,43570.43738,46210.46395,
48103.48214,48680.48900)
/gene="SREBF2"
/note="match: proteins: Sw:Q12772 Sw:Q60429 Sw:P36956
Sw:Q60416 Tr:O15352"
/codon_start=3
/evidence=not experimental
/product="bK25OD10.1 (sterol regulatory element binding
transcription factor 2)"
/protein_id="CA84134.1"
/db_xref="G1:5578715"
/training="ML04P5N0VGBRDLFSHOLCSSPFGSGSGSSSGSSSSS
NGRSGSSGSSVDSVORSFOVTLPPSPSPAPRPOATLONATSPSPVPTPAPLQ
PRGPOPOPOLOOQVATLPPSTSTPAPRPOLOOATONATSPSPVPOQSVLVT
SSONOPVTLCOQVTOVQORVLTORANGLOTLPATVPAPOVQVTVVOPOLI
KTDSVTLTKKTSBPVMAVONPALTLTTLTQALDQVTLVSGSGLTTLTPPM
GQKRPVQVPGSGVQOLBPBKGKERTTN11EKRRSS1NK11ELKOLVGTDAK
HNSGVLRAKIDATYLOQVNGH1RQENW1K1ANONK1K1GIDSLVADNED1K1B
DEKQVLA1RA1SPADISQVAGSP1SIDBSPSLDDAKVDEPSPVLA1KVD1R
R1LCV1P1C1SFPD1T1L1MGAGHSD1DHP1SGSGSV1SPFSGSGGFWPMWPT
LL1M1V1Y1V1SVFK1L1VHBP1VR1P1R1S1R1S1Y1P1WR1KQAD1DL1AGD1PAA1N
LQ1T1A1V1R1A1P1S1R1D1A1G1S1N1V1R1Y1S1Q1K1L1V1W1K1R1Q1C1R1A1P1A1G
FEDK1T1A1R1A1D1A1V1R1H1R1H1A1T1G1K1P1A1G1A1C1SD1V1N1A1C1A1V1A1E1K1P1P1
V1E1H1T1A1M1K1R1C1G1G1A1T1F1G1A1S1T1F1R1A1S1L1C1P1H1S1A1V1D1S1R1W1C1G1Q1K1P1P1

```

Alignment Scores:		
Pred. No.:	3.02e-25	Length: 220895
Score:	765.00	Matches: 138
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	59.81%	Indels: 0
Dbs:	9	Gaps: 0

[illegible]

QY	158	GI:Y1A1APProLeuG1Y1YrG1Y1A1APProLeuG1Y1YrG1Y1ThnProProLeuG1Y	177
Db	170185	GGAGCCCCCACTCTAGAGATATGAGACCCCACTCTTGAGATATGAAACCCCACTCTGGGA	170244
QY	178	TYrG1Y1A1APProLeuG1Y1YrG1Y1A1APProProLaG1YaaG1uG1YProProLa	197
Db	170245	TATGGAGCCCACTCTCTGGATATGAGAGCCCACTGAGGAAATGAAAGCCGCGCTGGCG	170304
QY	198	GI1Y1YrG1A1AserProLaG1YserG1Y1Aa1p1roG1uG1uSerTh1A1A1G1n	217
Db	170305	GGATTAACAAGAGCCCTCACTCTGATACAGAGAGCAGAGCTTCAGGAATCAAGACAGCCAG	170364
QY	218	AlAProGluAaGluA1AserA1uSerA1AserA1AserA1A1H1A1Ser	235
Db	170365	GCTCCTGAAACAGAGCTTCTCTCTCCCTCTGCTCTCTCTCTCAAGTCCATCTT	170418
RESULT 6			
LOCUS	AC113593		
DEFINITION	mus musculus clone RP23-363124, WORKING DRAFT SEQUENCE, 18 ordered		
ACCESSION	AC113593	168425 bp	DNA linear HTG 06-JUN-2002
VERSION	AC113593.3	GI:21327432	
KEYWORDS	HTG, HTGS PHASE2, HTGS DRAFT, HTGS_FULLTOP.		
SOURCE	mus musculus (house mouse)		
ORGANISM	mus musculus		
REFERENCE	Bukacynska; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus; 1 (bases 1 to 168425)		
AUTHORS	Bitren,B., Linton,L., Nusbaum,C. and Lander,B.		
TITLE	Mus musculus, clone RP23-363124		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 168425)		
AUTHORS	Bitren,B., Linton,L., Nusbaum,C., Lander,B., Ali,A., Allen,N.,		

TITLE	JOURNAL
Direct Submission Submitted (03-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA 3 (pages 1 to 168425)	REFERENCE
Bitnen, B., Linton, L., Nisbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barina, N., Bastien, V., Bloom, T., Bogunlavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collimore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dode, S., Gao, S., Ferreira, P., Fitzgerald, M., Fitzhugh, M., Gage, D., Ferguson, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, J., LaRoque, K., Lamzas, R., Landers, T., Lechoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, S., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McErmann, K., Meldrum, J., Meneses, L., Mihova, T., Mienga, V., Murphy, T., Naylor, T., Nguyen, C., Nicot, R., Nordby, C., Norman, C.H., O'Connor, T., O'Donnell, P.,	REFERENCE AUTHORS

KEYWORDS HTG; HTGS PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE Sus scrofa (pig)

ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

AUTHORS 1 (bases 1 to 179222)

TITLE Tracey.A.

JOURNAL Direct Submission

Submitted (25-MAY-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

On May 26, 2003 this sequence version replaced gi:31043704.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquerry@sanger.ac.uk

----- Project Information

Center project name: bnf121D21

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 177502 bases at least Q40

Consensus quality: 177971 bases at least Q30

Consensus quality: 178283 bases at least Q20

Insert size: 178722; sum-of-contigs

Insert size: 167315; 12.1% error; agarose-fp

Quality coverage: 7.82x in Q20 bases; sum-of-contigs Quality coverage: 8.59x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 5850: contig of 5850 bp in length

* 5851 5950: gap of 100 bp

* 5951 74264: contig of 68314 bp in length

* 74265 74364: gap of 100 bp

* 74365 84656: contig of 10292 bp in length

* 84657 84756: gap of 100 bp

* 84757 108493: contig of 23737 bp in length

* 108494 108593: gap of 100 bp

* 108594 114454: contig of 5861 bp in length

* 114455 114554: gap of 100 bp

* 114555 179222: contig of 64668 bp in length.

FEATURES

source 1..179222

location/Qualifiers

1..179222

/organism="Sus scrofa"

/mol_type="genomic DNA"

/db_xref="taxon:9823"

/clone="pigR-121D21"

/clone_id="pigR"

1..5850

/note="assembly fragment:02838"

fragment_chain:1

5951..74264

/note="assembly fragment:01479"

fragment_chain:1

74365..84656

/note="assembly fragment:01113"

fragment_chain:1

84757..108493

/note="assembly fragment:00846"

fragment_chain:1

108594..114454

/note="assembly fragment:01058"

fragment_chain:2

114555..179222

/note="assembly fragment:02416"

BASE COUNT 44892 a 44880 c 44117 g 44831 t 502 others

ORIGIN

Alignment Scores:

Pred. No.: 7,83e-10 Length: 179222

Score: 424.00 Matches: 92

Percent Similarity: 50.53% Conservative: 3

Best Local Similarity: 48.94% Mismatches: 41

Query Match: 33.15% Indels: 52

DB: 2 Gaps: 5

US-09-864-291-12 (1-235) x BX296515 (1-179222)

QY 96 CysSerValIleValIYrGly----- 102

DB 55262 TGCCAGTGTGGCTATGAGCCCTCAGTAGATATGAGGCAACAGCAAGATAT 55321

QY 103 ---AlaProProAlaGlyYrGlyYAlaProProProGlyYrGlyYAlaProProAlaGly 121

DB 55322 AGAAGCCCACTCCAGATACGAGGCCCACTCCAGATACGAGGCCCACTCCAGAA 55381

QY 122 TyrGlyAlaGlnProValGlyAenGlyYrProProValGlyYrAraGAlaSerProVal 141

DB 55382 TAGGAGCCCACTCCAGATACGAGGCCCACTCCAGATACGAGGCCCACTCGAA 55441

QY 142 ArgTyrGlyAlaProProLeuGlyYrGlyYAlaPro----- 153

DB 55442 AGACTCGAAGCCCAAGAGGATATGAGGCCCACTCGGGATATGAAAGCTCACT 55501

QY 154 -----ProAlaGlyYrGlyY-----AlaPro 160

DB 55502 GCGAGATATGACGCTCCCTCCAGATATGAGCTCACTCAAGAAAGCAAGAGCCCA 55561

QY 161 ProLeuGlyYrGlyYAlaProProLeuGlyYrGlyYrProProLeuGlyYrGlyYAla 180

DB 55562 CCACAGATATGAGGCTCCCTCCAGATATGAGGCCCACTCGAAGATATGATGCC 55621

QY 181 ProProLeuGlyYrGlyYAlaProPro----- 189

DB 55622 CCGCTCCAGATATGAGGCTCACTCAAGAAAGATGAGGCCCTCGAGATATGAA 55681

QY 190 -----AlaGlyAenGlyYrProProAlaGlyYrAraGAlaSerProAlaGlySer 206

DB 55682 CACCAACAGCTGAAATGAAAGCCCACTCAAGATATGAAAGCCCACTCGTGAAT 55741

QY 207 GlyAlaArgProGlnGlnSerThrAlaGlnAlaProGlnGlnGlnAlaSerLeuPro 226

DB 55742 AGAGCTGCTCTCAACAAATCTGAGGAGCCCAAGCCG-----GAGGCTTCTTCCC 55792

QY 227 SerAlaSerSerGlnValHis 234

DB 55793 TCTACCTCATCTTCTCAGGCCCA 55816

RESULT 9

BX470149 222469 bp DNA linear HTG 07-MAY-2003

LOCUS BX470149

DEFINITION Danio rerio clone DKX-242K7, *** SEQUENCING IN PROGRESS ***, 50

unordered pieces.

ACCESSION BX470149

VERSION BX470149.3 GI:30424228

KEYWORDS HTG; HTGS PHASE1.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 222469)

AUTHORS Burton.J.

TITLE Direct Submission

JOURNAL Submitted (06-MAY-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk

COMMENT On May 7, 2003 this sequence version replaced gf:30387077.

----- Genome Center
Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zfish-help@sanger.ac.uk

----- Project Information

Center project name: ZK242K7

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 197958 bases at least Q40

Consensus quality: 206715 bases at least Q30

Consensus quality: 212351 bases at least Q20

Insert size: 217569; sum-of-contigs

Insert size: 165837; 6.2% error; agarose-fp

Quality coverage: 2.48x in Q20 bases; sum-of-contigs Quality coverage: 3.95x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1      6656: contig of 6656 bp in length
*      6657      6656: gap of 100 bp
*      6757      10255: contig of 3499 bp in length
*      10256      10355: gap of 100 bp
*      10356      13223: contig of 2868 bp in length
*      13224      13323: gap of 100 bp
*      13324      19935: contig of 6612 bp in length
*      19936      20035: gap of 100 bp
*      20036      22597: contig of 2562 bp in length
*      22598      22697: gap of 100 bp
*      22698      28569: contig of 5872 bp in length
*      28570      28669: gap of 100 bp
*      37965      37965: contig of 9296 bp in length
*      37966      38065: gap of 100 bp
*      38066      41140: contig of 3075 bp in length
*      41141      41240: gap of 100 bp
*      41241      45108: contig of 3868 bp in length
*      45109      45208: gap of 100 bp
*      45209      47730: contig of 2522 bp in length
*      47731      47830: gap of 100 bp
*      47831      51884: contig of 4054 bp in length
*      51885      51984: gap of 100 bp
*      51985      55857: contig of 3873 bp in length
*      55858      55957: gap of 100 bp
*      55959      60498: contig of 4541 bp in length
*      60499      60598: gap of 100 bp
*      60599      63641: contig of 3443 bp in length
*      63642      63741: gap of 100 bp
*      63742      72674: contig of 8933 bp in length
*      72675      72774: gap of 100 bp
*      72775      75602: contig of 2828 bp in length
*      75603      75702: gap of 100 bp
*      75703      78062: contig of 2260 bp in length
*      78063      78162: gap of 100 bp
*      78163      84485: contig of 6323 bp in length
*      84486      84585: gap of 100 bp
*      84586      87553: contig of 2968 bp in length
*      87554      87653: gap of 100 bp
*      87654      92487: contig of 4834 bp in length
*      92488      92587: gap of 100 bp
*      92588      96104: contig of 3517 bp in length
*      96105      96204: gap of 100 bp
*      96205      101479: contig of 5275 bp in length
*      101480      101579: gap of 100 bp
*      101580      104979: contig of 3400 bp in length
*      104980      105079: gap of 100 bp
```

```
*      105080      108003: contig of 2924 bp in length
*      108004      108103: gap of 100 bp
*      108104      113151: contig of 5048 bp in length
*      113152      113251: gap of 100 bp
*      113252      116585: contig of 3334 bp in length
*      116586      116685: gap of 100 bp
*      116686      119346: contig of 2661 bp in length
*      119347      119446: gap of 100 bp
*      119447      123125: contig of 3679 bp in length
*      123126      123225: gap of 100 bp
*      123226      131466: contig of 8041 bp in length
*      131467      131667: gap of 100 bp
*      131668      136193: contig of 4827 bp in length
*      136194      136293: gap of 100 bp
*      136294      139482: contig of 3189 bp in length
*      139483      139582: gap of 100 bp
*      139583      141875: contig of 2233 bp in length
*      141876      141975: gap of 100 bp
*      141976      144122: contig of 2147 bp in length
*      144123      144222: gap of 100 bp
*      144223      146968: contig of 2746 bp in length
*      146969      147068: gap of 100 bp
*      147069      149399: contig of 2331 bp in length
*      149400      149499: gap of 100 bp
*      149500      153963: contig of 4464 bp in length
*      153964      154063: gap of 100 bp
*      154064      156579: contig of 2516 bp in length
*      156580      156679: gap of 100 bp
*      156680      158928: contig of 2249 bp in length
*      158929      159028: gap of 100 bp
*      159029      164542: contig of 5514 bp in length
*      164543      164642: gap of 100 bp
*      164643      172054: contig of 7412 bp in length
*      172055      172154: gap of 100 bp
*      172155      175210: contig of 3056 bp in length
*      175211      175310: gap of 100 bp
*      175311      179379: contig of 4069 bp in length
*      179380      179479: gap of 100 bp
*      179480      182698: contig of 3219 bp in length
*      182699      182798: gap of 100 bp
*      182799      186779: contig of 3981 bp in length
*      186780      186879: gap of 100 bp
*      186880      195626: contig of 8747 bp in length
*      195627      195726: gap of 100 bp
*      195727      199386: contig of 3670 bp in length
*      199387      199496: gap of 100 bp
*      199497      206412: contig of 6916 bp in length
*      206413      209619: contig of 3107 bp in length
*      209620      209719: gap of 100 bp
*      209720      219083: contig of 9364 bp in length
*      219084      219183: gap of 100 bp
*      219184      222469: contig of 3286 bp in length.
```

FEATURES

source

```
1..222469
  /organism="Danio rerio"
  /mol_type="genomic DNA"
  /db_xref="taxon:7955"
  /clone="DKFZ-242K7"
  /clone_idb="DantioKey"
```

misc_feature

```
1..6656
  /note="assembly fragment:01359
  fragment_chain:1"
  /note="assembly fragment:01385
  fragment_chain:1"
```

misc_feature

```
10356..13223
  /note="assembly fragment:00592
  fragment_chain:1"
  /note="assembly fragment:00237
  fragment_chain:1"
```

misc_feature

```
20036..22597
```



```

/note="clone
clone end:T7

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sotacungathi; Muridae; Murinae.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

FEATURES
source
location/Qualifiers
1..129624
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9623"
/clone="PigB-211K18"
/clone_1lb="PigB"
1..4262
/note="assembly_fragment:00361
fragment_chain:1"
misc_feature
4363..6991
/note="assembly_fragment:00108
fragment_chain:1"
7092..9480
/note="assembly_fragment:00460
fragment_chain:1"
9581..11688
/note="assembly_fragment:00959
fragment_chain:1"
11789..14658
/note="assembly_fragment:00995
fragment_chain:2"
14759..18440
/note="assembly_fragment:00608
fragment_chain:2"
18541..20818
/note="assembly_fragment:01101
fragment_chain:2"
20919..22969
/note="assembly_fragment:00996
fragment_chain:2"
23070..27584
/note="assembly_fragment:00217
fragment_chain:3"
27685..33211
/note="assembly_fragment:00019
fragment_chain:3"
33312..35521
/note="assembly_fragment:00548
fragment_chain:3"
35622..38528
/note="assembly_fragment:00815
fragment_chain:4"
38629..42063
/note="assembly_fragment:00246
fragment_chain:4"
42164..44164
/note="assembly_fragment:00859
fragment_chain:5"
44265..47707
/note="assembly_fragment:00964
fragment_chain:5"
47808..50737
/note="assembly_fragment:01192
fragment_chain:6"
50838..57066
/note="assembly_fragment:00542
fragment_chain:6"
57167..59978
/note="assembly_fragment:01304
fragment_chain:7"
60079..62560
/note="assembly_fragment:00948
fragment_chain:7"
62661..64963
/note="assembly_fragment:00016"
65064..67191
/note="assembly_fragment:00065"
67292..72005
/note="assembly_fragment:00070"
72106..76595
/note="assembly_fragment:00136"
76696..79825
misc_feature

misc_feature
79926..83540
/note="assembly_fragment:00162"
79926..83540
/note="assembly_fragment:00367"
83641..86276
/note="assembly_fragment:00492"
86377..94620
/note="assembly_fragment:00545"
94721..97520
/note="assembly_fragment:00745"
97621..99784
/note="assembly_fragment:00801"
misc_feature
Alignment Scores:
Pred. No.: 4.38e-08 Length: 129624
Score: 383.00 Matches: 76
Percent Similarity: 57.66% Conservative: 3
Best Local Similarity: 55.47% Mismatches: 44
Query Match: 29.95% Indels: 14
Gaps: 1
US-09-864-291-12 (1-235) x BX236540 (1-129624)
QY 96 CyservallieVallyTyrGly----- 102
DB 65506 TGCCCACTTGTGCTCTATGACCCCTACAGTAGATATGAGGCCAACAGCAAAATAT 65447
QY 103 ---AlaProProAlaGlyTyrGlyValaProProProGlyTyrGlyValaProAlaGly 121
DB 65446 AGAGCCCACTCCAGATATGAGAGCCCACTCCAGATATGAGAGCCCACTCCAGAG 65387
QY 122 TyrGlyAlaGlnProValaGlyValaGlnGlyProProValaGlyTyrGlyAlaSerProVal 141
DB 65386 TACGAGGCCCACTCCAGATATGAGAGCCCACTCCAGATATGAGAGCCCACTCA 65327
QY 142 ArgTyrGlyAlaProProLeuGlyTyrGlyValaProProAlaGlyTyrGlyValaProPro 161
DB 65326 AGACTCGAAGCCCACTCCAGATATGAGAGCCCACTCCAGATATGAGAGCCCACT 65267
QY 162 LeuGlyTyrGlyValaProProLeuGlyTyrGlyValaProProLeuGlyTyrGlyValaPro 181
DB 65266 GCGAGATATGAGAGCCCTCCCTCCAGATATGAGAGCCCTCCAGATATGAGAGCCCA 65207
QY 182 ProLeuGlyTyrGlyValaProProAlaGlyValaGlnGlyProProAlaGlyTyrGlyAla 201
DB 65206 CCAAGAGATATGAGAGCCCACTCCAGATATGAGAGCCCACTCCAGATATGAGAGCC 65147
QY 202 SerProAlaGlySerGlyValaArgProGlnuSerThrAlaAlaGlnAla 218
DB 65146 CCGCTCCAGATATGAGAGCCCTCCAGATATGAGAGCCCTCCAGATATGAGAGCC 65096
RESULT 13
AC092212 170781 bp DNA linear HTG 05-APR-2002
LOCUS AC092212
DEFINITION Trypanosoma brucei chromosome VIII clone RPI393-10J17, ***
SEQUENCING IN PROGRESS ***
AC092212
AC092212.31 GI:20043109
VERSION HTG; HTGS_PHASE2.
KEYWORDS Trypanosoma brucei
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma
Bukariola; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma
1 (bases 1 to 170781)
El-Sayed,N.M., Ghedin,B., Song,J., Larkin,C., Wanless,D., Jones,K.,
Peterson,J., Hou,L., Zhao,H., Mason,T., Miltchev,J., Pal,G., Van
Aken,S., Utecherback,T., Khalak,H.G., Gerard,C., Leech,V., Ullu,B.,
Melville,S., White,O., Adams,M.D., Donaldson,J.B. and Fraser,C.M.
Trypanosoma brucei GUTc10.1 RPI393-10J17 BAC genomic sequence
Unpublished
2 (bases 1 to 170781)
El-Sayed,N.M., Khalak,H. and Adams,M.D.
Direct Submission
JOURNAL
AUTHORS
TITLE
JOURNAL
Submitted (29-JUN-2001) The Institute for Genomic Research, 9712

REFERENCE 3 (bases 1 to 170781)
 Medical Center Dr, Rockville, MD 20850, USA
 AUTHORS El-Sayed, N.M., Khalak, H. and Adams, M.D.
 TITLE Direct Submission
 JOURNAL Submitted (05-APR-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 COMMENT On Apr 5, 2002 this sequence version replaced gi:18425287.
 * NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
 * This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
 * 1 170781: contig of 170781 bp in length.
 FEATURES
 source
 1..170781
 /organism="Trypanosoma brucei"
 /mol_type="genomic DNA"
 /isolate="GUTat10.1"
 /db_xref="taxon:5691"
 /chromosome="VIII"
 /clone="RPC193-10J17"
 BASE COUNT 42610 a 38223 c 43377 g 46571 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.00102 Length: 170781
 Score: 289.00 Matches: 73
 Percent Similarity: 47.50% Conservative: 3
 Best Local Similarity: 45.62% Mismatches: 50
 Query Match: 22.60% Indels: 34
 Gaps: 11
 DB: 2
 US-09-864-291-12 (1-235) x AC092212 (1-170781)
 QY 87 GYAsmNeCythrProglMet-----ProCySsrvalIleval 100
 Db 65879 GGGGGGAGTGTTCACGAGCTTAATGTGCTGGCTGGGCTCCACAGATAATGAGAAA 65938
 QY 101 TrrGlyAlaProProAlaGlyTrrGlyAlaProProPro-----GlyTrrGly----- 116
 Db 65939 CAGCCCCCACCACCTGCGGATATGTCAACCCACACCTGCGGATATGTCAACCC 65998
 QY 117 AlaProAlaGlyTrrGlyAlaGlyAlaProValGlyAsnGlyProProValGlyTrr 136
 Db 65999 CCACCACTGCGGATATGTCAACCCCA-----CCACCTGCGGATAT 66043
 QY 137 ArgAlaSerProValArgTrrGlyAlaProProLeuGlyTrrGly-----AlaProPro 154
 Db 66044 GGTGAGCCCCA-----CCACCTGCGGATATGTCAACCCCAACCACT 66088
 QY 155 AlaGlyTrrGly-----AlaProProLeuGlyTrrGly-----AlaProProLeuGly 170
 Db 66089 GCCCGATATGTCAACCCCAACCACTGCGGATATGTCAACCCCAACCACTGCGGGA 66148
 QY 171 TrrGly-----ThrProProLeuGlyTrrGly-----AlaProProLeuGlyTrrGly 186
 Db 66149 TATGTGAGCCCCCAACCACTGCGGATATGTCAACCCCAACCACTGCGGATATGT 66208
 QY 187 AlaProProAla-----GlyAsnGlyProProAlaGlyTrrGlyAlaSer 202
 Db 66209 CAGCCCCCACCACCTGCGGATATGTCAACCCCAACCACTGCGGATATGTCAACCG 66268
 QY 203 ProAlaGlySerGlyAlaArgProGlnGlnSerThrAlaAlaGlnAlaProGlnAsnGln 222
 Db 66269 CCGTGTGCGCTTCAACAACCACTGCGGATATGTCTGAGTCAAAAGTGGAGAGGAG 66328
 RESULT 14
 LOCUS AB072784 2170 bp mRNA linear PRI 22-FEB-2003

DEFINITION Macaca fascicularis testis cDNA clone:QcSA-20807, full insert sequence.
 ACCESSION AB072784
 VERSION AB072784.1 GI:16041171
 KEYWORDS oligo capping; f1s (full insert sequence).
 SOURCE Macaca fascicularis (craab-eating macaque)
 ORGANISM Macaca fascicularis
 Bacteria; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butleria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.
 REFERENCE 1
 Oseada, N., Hida, M., Kusuda, J., Tanuma, R., Hirata, M., Suto, Y., Hirai, M., Teruo, K., Sugano, S. and Hashimoto, K.
 Cynomolgus monkey testicular cDNAs for discovery of novel human genes in the human genome sequence
 BMC Genomics 3 (1), 36 (2002)
 JOURNAL PUBMED 12498619
 REFERENCE 2 (bases 1 to 2170)
 Hashimoto, K., Oseada, N., Hida, M., Kusuda, J. and Sugano, S.
 Direct Submission
 JOURNAL Submitted (09-OCT-2001) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
 (E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genbank/, Tel: 81-3-5285-1111 (ex.2120), Fax: 81-3-5285-1181)
 Lab host: TOP10
 Vector: pME18S-FL3 (Acc.No. AB090864)
 R. Site1: DraIII (CACTGATG)
 R. Site2: DraIII (CACCATGTG)
 Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGAGCCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by oligo-capping method (Sugano et al., University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' end primer [CTTCTGCTTAAAGCTGCG]; 3' end primer [CGACCTGACGCTCAAGCAC])
 Location/Qualifiers
 1..2170
 /organism="Macaca fascicularis"
 /mol_type="mRNA"
 /db_xref="taxon:9541"
 /clone="QcSA-20807"
 /sex="male"
 /tissue_type="testis"
 /clone_id="macaque testis cDNA library QcSA"
 /dev_stage="adult"
 /dev_1171
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="BAB69753.1"
 /db_xref="GI:16041172"
 /translation="MFPNCSGSLMSYDVELTENDKNEVBAFRTKGTGYLTPT RYFLSKGDAMQSPMPPYLMKDCIKQPVGSANTIKGVAAAGGSGWGSAYLT FTAGALIFEGQNLGVASQASRGALPAGNAGYVPSGAYVPPVPAANGYPCPPGP YPPPPPPYRPPMDGAMGAYGPPPPYRGMPPVPSGPDVPSPTAABAKAABAAAS ATNPNPNPNVNPPTSQPPPPYPPEDKKTQ"
 BASE COUNT 464 a 663 c 544 g 499 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.00111 Length: 2170
 Score: 276.50 Matches: 90
 Percent Similarity: 38.49% Conservative: 22
 Best Local Similarity: 30.93% Mismatches: 76
 Query Match: 21.62% Indels: 103
 Gaps: 12
 DB: 9
 US-09-864-291-12 (1-235) x AB072784 (1-2170)

Oy		1	MctProPheAapLeuMetThrAsnLeuThrValGluGlnProValPheAlaIAspHis	20
Db		599	ATGCCGTTTATCTCATTAAGAACTGTGAATCAACGACCCTGATTGGTCGAACACTAC	658
Oy		21	IleArglyThrIleGlnAlaLaProTyrglyTYTPGluGylGlnAlaThrPheLys	40
Db		659	ATCAAAGGAACAGTGAAAGCGGAACGGGAAGGTGACTGGAAAGGCTGCTTCCACAAAG	718
Oy		41	LeuValPheAlaGlaanglyAspAlallegluPhealaglnLeuMetValLysAlaIAser	60
Db		719	TTCACCTTACAGGCAGGGGGCGCCATTGAATTGACAAGCGGAATCTCAGATGGCACT	778
Oy		61	AlaValAlaArglyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIle	80
Db		779	CAGCCTCAAGGT-----GAAAGCCCCAATGAGCC	811
Oy		81	TyrValIlethrnglygluglyAsnMetCysThrProGlnMetProCysSerValIleVal	100
Db		812	TATGGCTACTCT-----TACATGGCCAGGGGGCTGATGTC	847
Oy		101	TyrGlyAlaProProAlaGlyTYrtyrAlaPro--ProProGlyTYrtyrAlaProPro	119
Db		848	TATCCCCCGCACAGTCGCCAATGAATGATACCCCTGACCTCTGGCTACCCCTATCCACCG	907
Oy		120	AlaGlyTYrtyrAlaGlnProValGlyAngluGlyProProValGlyTYrtyrAlaIAser	139
Db		908	CCC-----CCGCTGAGTTCTATCCAGAACC	934
Oy		140	ProValArGTYrtyrAla-----ProProLeuGlyTYrtyrAlaPro	153
Db		935	CCCATGATGACGGGGGCATGGAGATACGTGCAGCCCCCACCACACCGCCTATCTGGGCC	994
Oy		153	-----	153
Db		995	ATGAACCTCGGCTCAGCGGCCCGCATGTCCCTTCACTCTGCAGCCGAAGCCAAGGCC	1054
Oy		154	-----Pro	154
Db		1055	GCAAGAGACGCCGCCAGGGCCTACTACACCAACCGAACCTCAACAAGCTACATGACC	1114
Oy		155	AlaGlyTYrtyrAlaProProLeuGlyTYrtyl-	165
Db		1115	ACGAGCCAGCGCCACACTCTCTACTACCAACCGGAAGATAAGAACCCAGTAGGCC	1174
Oy		166	-----AlaProProLeuGly---TYrtylTHrProProLeuGlyTYrtyl	179
Db		1175	CTCCTGCTCCCTGCTGCTCCACCTTATCTCTGATACCTTACCTCCCTCCCATGGG---GCT	1231
Oy		180	AlaProProLeuGlyTYrtyrAlaProProAlaGlyAngluGlyProProAlaGlyTYr	199
Db		1232	GCACTGGGGGCTGGGGGAGGGGGCGCTTGTCTCCCTCAG-----TCTGATCAT	1287
Oy		200	ArgAlaSerProAlaGlySer-----GlyAlaArgProGlnGlu-----	212
Db		1283	AACACATTACCAAGAACTAGCATTTGTGGGAATTAGGGCCCCCGGCTCAAGAGAGTGC	1342
Oy		213	-----SerThrAlaAlaGlnAlaProGlu	220
Db		1343	CGCCAGCTTCCATGCCAGCCCAAGCCCAACAA	1375
RESULT 15				
LOCUS	AC134957	166045 bp	DNA	linear HTG 23-NOV-2002
DEFINITION	Tetradodon nigroviridis clone G5MB-4505, WORKING DRAFT SEQUENCE,	3		
ORDERED	pieces.			
VERSION	AC134957			
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT.			
SOURCE	Tetradodon nigroviridis			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;			
	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;			

REFERENCE	1 (bases 1 to 166045)
AUTHORS	Alberich N., Antoncollis A., Ayele R., Beckstrom-Sternberg S.M., Benjamin B., Blakeley R.W., Bouffard G.G., Brinley C., Brooks S., Carlsaga K., Coleman B., Engle J., Granite S., Guan X., Gupta V., Haghighi P., Han J., Hansen N., Ho S.-L., Idol J.R., Kallins B., Kauric P., Lee-Lin S.-Q., Legaspi R., Maduro Q.L., Maduro V.B., Maricello B.H., Mastello C., Masekri B., McDowell J., Padurigan C., Pearson R., Portnoy M.E., Prasad A., Reddy-Dugue N., Schandler K., Schnuler M.G., Sison C., Stancirp S., Thomas J.W., Thomas P.J., Touchman J.W., Vogt J.L., Wecherby K.D., Wiggins L., Young A. and Green E.D.
TITLE	NISC Comparative Sequencing Initiative
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 166045)
AUTHORS	Green E.D.
TITLE	Direct Submission
JOURNAL	Submitted (03-OCT-2002) NIH Intramural Sequencing Center, 8717 Groveomont Circle, Galthersburg, MD 20877, USA
REFERENCE	3 (bases 1 to 166045)
AUTHORS	Green E.D.
TITLE	Direct Submission
JOURNAL	Submitted (23-NOV-2002) NIH Intramural Sequencing Center, 8717 Groveomont Circle, Galthersburg, MD 20877, USA
COMMENT	On Nov 23, 2002 this sequence version replaced gi:23477837.

----- Genome Center -----
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc_zoo@hgt.nih.gov
Project Information
Center project name: dgd
Center clone name: 045005

The sequence data in this record represents an 'enhanced' version of a phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165749 bases at least Q40
Consensus quality: 165822 bases at least Q30
Consensus quality: 165836 bases at least Q20
Insert size: 175000; agarose-fp
Insert size: 165865; sum-of-contigs
Quality coverage: 13.18x in Q20 bases; agarose-fp
Quality coverage: 13.91x in Q20 bases; sum-of-contigs

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of 'N'. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 44384: contig of 44384 bp in length
* 44385 44484: gap of unknown length
* 44485 55090: contig of 10606 bp in length
* 55091 55191: gap of unknown length
* 55191 166045: contig of 110855 bp in length.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - nucleic search, using frame_plus.p2n model

Run on: December 16, 2003, 11:00:59 ; Search time 196.834 Seconds

(without alignments)
3222.858 Million cell updates/sec

Title: US-09-864-291-12

Sequence: 1 MPEDLMTNLTVEQVFAANF.....AQAPEKASLPASSSQVHS 235

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1345719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ .p2n.model -DRV=xlp
-O=/cgm2.1/USPTO.spool.p/US09864291/runat.15122003.160858.23296/app.query.fasta_1.846
-DB=N Geneseq.19Jun03 -OPMT=fastcap -SUFFIX=ring -MIMATCH=0.1 -LOOPTC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdt
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTMT=ptco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09864291.OCGN.1.1.511.0runat.15122003.160858.23296 -NCPU=6 -ICPU=3
-NO MAMP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -MAIN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq.19Jun03.*
1: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1279	100.0	1001	24	AAS20602	DNA encoding human
2	826	64.6	467	22	ABL01229	Human reproductive
3	826	64.6	467	23	ABL96688	Human testicular a
4	777	60.8	1413	24	AAS20601	DNA encoding bovin
5	765	59.8	220895	24	ABK84798	Human cDNA differe
6	761	59.5	436	22	ABA67850	Human foetal liver
7	761	59.5	436	22	AAK42003	Human bone marrow
8	761	59.5	436	22	AAI48070	Probe #16756 used
9	761	59.5	436	24	ABSI6034	Human genome-deriv
10	705	55.1	7099	22	AAI04882	Human reproductive
11	705	55.1	7099	23	ABJ79776	Human testicular a
12	342	26.7	894	23	AAS70582	DNA encoding novel
13	261.5	20.4	1885	24	ABJ70430	Human bone remodel
14	258	20.2	1915	21	AAJ22363	Human secreted pro
15	258	20.2	1915	25	ABJ73614	Human secreted pro
16	258	20.2	1915	25	ABJ73614	Human secreted pro
17	258	20.2	1915	25	ABJ73614	Human secreted pro
18	252.5	19.7	1109	23	ABJ17151	Drosophila melanog
19	245	19.2	471	22	ABA55746	Human foetal liver
20	245	19.2	471	22	AAK28957	Human bone marrow
21	245	19.2	471	22	AAI34906	Probe #3592 used t
22	245	19.2	471	24	ABJ03491	Human genome-deriv
23	220.5	17.2	4881	19	AAV63438	Feline herpesvirus
24	215	16.8	2614	23	ABL28430	Drosophila melanog
25	215	16.8	2614	23	ABL05187	Drosophila melanog
26	215	16.8	2614	23	ABL05186	Drosophila melanog
27	209.5	16.4	756	17	AAV16766	Drosophila-like poly
28	206.5	16.1	756	14	AAQ403032	Collagen-like poly
29	204.5	16.0	4329	23	ABL28034	Drosophila-like poly
30	202.5	15.8	1863	23	ABL28035	Drosophila melanog
31	197.5	15.4	4323	23	ABL17150	Drosophila melanog
32	195	15.2	432	17	AAV16765	Collagen-like poly
33	194	15.2	468	23	ABL28431	Drosophila melanog
34	192.5	15.1	2853	23	ABL29393	Drosophila melanog
35	192.5	15.1	6906	23	ABL29392	Drosophila melanog
36	192	15.0	321	22	AAI01355	Human reproductive
37	192	15.0	321	23	ABL96808	Human testicular a
38	192	15.0	4403765	22	AAI99683	Mycobacterium tube
39	192	15.0	4411529	22	AAI99682	Mycobacterium tube
40	191	14.9	758	22	AAI23504	Human breast cancer
41	189.5	14.8	3206	23	ABL16637	Drosophila melanog
42	189.5	14.8	5595	23	ABL16636	Drosophila melanog
43	188.5	14.7	432	14	AAQ43031	Collagen-like poly
44	186	14.5	2144	18	AAV69165	Trypanosoma cruzi
45	186	14.5	2144	20	AAK81754	DNA encoding a T.

ALIGNMENTS

RESULT 1

AAS20602
AAS20602 standard; cDNA, 1001 BP.

AC AAS20602;

DT 09-APR-2002 (first entry)

DB DNA encoding human testicular WW domain binding protein (htwbp).

XX Testicular WW domain binding protein; WBPI; perinuclear theca 32;

KW P732; contraceptive; fertility; oocyte activation; vaccine;

KW globozoosperm; spermatogenesis; spermatocyst; tyrosine kinase; C-Yes;

XX Immunoreceptor; human; gene; ss.

XX Homo sapiens.

```

FH Key Location/Qualifiers
FT primer_bind 1..18
FT /tag= a
FT /note= "Binds forward primer AAS20605"
FT CDS 1..708
FT /tag= b
FT /product= "htwbp"
FT /note= "Human testicular WW domain binding protein"
FT primer_bind complement (984..1001)
FT /tag= c
FT /note= "Binds reverse primer AAS20606"
XX
XX WO200190185-A2.
XX
XX PD 29-NOV-2001.
XX
XX PF 25-MAY-2001; 2001WO-CA00738.
XX
XX PR 25-MAY-2000; 2000CA-2307128.
XX
XX PR 25-MAY-2000; 2000US-206979P.
XX
XX PA (TOOH ) UNIV QUEENS KINGSTON.
XX PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX PI Oki R, Sutovsky P;
XX
XX WIPI, 2002-097644/13.
XX
XX DR P-PSDB; AAU74610.
XX
XX PT Isolated perinuclear theca 32 polypeptide that interacts with activated
XX PT tyrosine kinase c-yes, for enhancing fertility, treating/diagnosing
XX PT diminished fertility and abnormal spermiogenesis and for providing
XX PT contraception -
XX
XX PS Claim 62; Fig 4B; 103pp; English.
XX
XX CC The invention describes an isolated perinuclear theca 32 (PT32)
XX CC polypeptide (1) which interacts with tyrosine kinase c-yes. (1) is
XX CC useful for: enhancing fertility in a mammal; treating globozoosperm, by
XX CC expressing (1) in spermatozoa; inhibiting fertilisation, by introducing
XX CC (1) or its antigenic fragment into a mammal to elicit an immune
XX CC response; enhancing the ability of round spermatids to activate oocytes;
XX CC treating or diagnosing diminished fertility and abnormal spermiogenesis;
XX CC in providing contraception; identifying contraceptive and
XX CC fertility-enhancing agents. The polynucleotide is useful for producing
XX CC (1) by recombinant techniques, as vaccine, as diagnostic reagents, and
XX CC for chromosome identification. An antibody against (1) is useful in
XX CC immunological assays, in immunocontraceptive methods, to identify cells
XX CC expressing (1), and to purify (1) by affinity chromatography. A
XX CC transgenic animal is useful as an animal model for studying human
XX CC fertility and reproductive biology, and for screening compounds to
XX CC identify modulators of oocyte activation. The use of (1) prevents the
XX CC entry of components which are detrimental to embryonic development into
XX CC the oocyte during oocyte activation with crude sperm extract and avoids
XX CC the propagation of viruses such as HIV (human immunodeficiency virus) and
XX CC SIV (Saimian immunodeficiency virus) carried in the sperm. This sequence
XX CC encodes the human testicular WW domain binding protein (htwbp), described
XX CC in the method of the invention.
XX
XX SQ Sequence 1001 BP; 261 A; 254 C; 248 G; 238 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 2,58e-58 Length: 1001
XX Score: 1279.00 Matches: 235
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-09-864-291-12 (1-235) x AAS20602 (1-1001)
XX
XX 1 MetProPhaApLeuMetThrAnuLeuThrValGluGlnProValPheAlaIaIaPhe 20
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||

```

```

DB 1 ATGCCATTGATCTGATGACGAACCTCAGTGTGAACAACAGATTTGCTGAACTTC 60
QY 21 ILelysQLYThr11IegInAlaAlaProTYrGlyGlyTYrGlyGlyGlnAlaThrPheLys 40
DB 61 ATTAAGGAACTATTCCATGCGAGCTCCATATGCTGGGAGAGCAAGCACTTCTTAA 120
QY 41 LeuValPheArGAGAGGlyAaPaAlaIleGluPheAlaGlnLeuMetValLysAlaIaSer 60
DB 121 TTACTCTTCAGAAATGAGAGATGCAATTAATTTGCCAGATTGATGGTGAAGCTGCTCT 180
QY 61 AlaValAlaArGlyPheProLeuArGThrLeuAlaAspTrpPheSerSerMetGlyIle 80
DB 181 GCTGTGCCGAGAGATTTCACCTTAGAACCTTAATACCTGCTTCACTCATGGGAATT 240
QY 81 TyrValIleThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 100
DB 241 TATGTAATTAATCTGGGAGAGGAAATATATGATGCTCAGAGATGCTTCTTCAAGTATGTC 300
QY 101 TYrGlyAlaProProAlaGlyTYrGlyAlaProProProGlyTYrGlyAlaProProAla 120
DB 301 TATGGAGCCCACTGAGAGATGAGAGCCCACTCCGAGATGAGAGCCCACTGCA 360
QY 121 GlyTYrGlyAlaGlnProValGlyAlaGlnGlyProProValGlyTYrAlaAlaSerPro 140
DB 361 GGATATGAGGCCCAACCGTAGAAATGAAGGCCCGCTGTGGATACAGAGCTCACT 420
QY 141 ValArGlyTYrGlyAlaProProLeuGlyTYrGlyAlaProProAlaGlyTYrGlyAlaPro 160
DB 421 GTGGAGATATGAGGCCCACTCTTGATATACAGAGCCCACTGAGATATGAGGCCCA 480
QY 161 ProLeuGlyTYrGlyAlaProProLeuGlyTYrGlyAlaProProLeuGlyTYrGlyAla 180
DB 481 CCTTAGAGATATGAGGCCCACTCTTGATATGAGAGCCCACTCTGAGATATGAGGCC 540
QY 181 ProProLeuGlyTYrGlyAlaProProAlaGlyAlaGlnGlyProProAlaGlyTYrArG 200
DB 541 CCACCTCTCGAGATATGAGGCCCACTGAGAGAAATGAAGCCCGCTGCGAGATACGA 600
QY 201 AlaSerProAlaGlySerGlyAlaArGProGlnGlnSerThrAlaAlaGlnAlaProGln 220
DB 601 GCTTCACCTGATGATGAGAGCCCACTGAGAGATCTACAGAGAGCCCAAGCTCTGAA 660
QY 221 AsnGluAlaSerLeuProSerAlaSerSerSerGlnValHisSer 235
DB 661 AACAGAGCTTCTCTTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
XX
XX RESULT 2
XX AAL01229
XX ID AAL01229 standard; cDNA, 467 BP.
XX
XX AC AAL01229;
XX
XX XX 21-NOV-2001 (first entry)
XX
XX DT Human reproductive system related antigen cDNA SEQ ID NO: 1230.
XX
XX DB Human reproductive system related antigen cDNA SEQ ID NO: 1230.
XX
XX KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ss.
XX
XX OS Homo sapiens.
XX
XX XX WO200155320-A2.
XX
XX PN 02-AUG-2001.
XX
XX PD 17-JAN-2001; 2001WO-US01339.
XX
XX PF 31-JAN-2000; 2000US-0179065.
XX
XX PR 04-FEB-2000; 2000US-0180628.
XX
XX PR 24-FEB-2000; 2000US-0184664.
XX
XX PR 02-MAR-2000; 2000US-0186350.
XX
XX PR 16-MAR-2000; 2000US-0189874.
XX
XX PR 17-MAR-2000; 2000US-0190076.

```

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220966.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0235802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
DR P-PSDB; AAM95259.
XX
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 1; SEQ ID NO 1230; 1297bp + Sequence listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC

CC including cancer. The present sequence is a coding sequence of the
invention.

XX Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;

Alignment Scores:

Pred. No.:	3,68e-35	Length:	467
Score:	826.00	Matches:	149
Percent Similarity:	96.77%	Conservative:	1
Best Local Similarity:	96.13%	Mismatches:	4
Query Match:	64.58%	Indels:	1
DB:	22	Gaps:	0

US-09-864-291-12 (1-235) x AAL01229 (1-467)

```
QY 44 ArgAsnGlyAspAlaIleGluPheAlaGlnLeuMetValIysAlaSerAlaValAla 63
DB 3 AGAAATGAGAGTGCATTTGTCAGTTGATGATGAAAGCTGCTGCTGCTGCC 62
QY 64 ArgGlyPheProLeuArgThrIleuAsnAspTrpPheSerSerMetGlyIleTyrValIle 83
DB 63 CGAGGATTTCCACTTAGAACCTTAATGACCTGGTTCAGTCTTAAGGAATTTATGTAAAT 122
QY 84 ThrGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleValTyrGlyAla 103
DB 123 ACTGGGGAAGGAAATATGTCACCTCCACAGATGCTGTTCAGTTATGCTAATGGAGCC 182
QY 104 ProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAlaProProAlaGlyTyrGly 123
DB 183 CCACCTGACGAGATATGAGGCCGCCACCTCCGAGATACGAGGCCGCCACCTGAGATATGGA 242
QY 124 AlaGlnProValGlyAsnGluGlyProProValGlyTyrArgAlaSerProValArgTyr 143
DB 243 GCCCAACCCGTAGGAATATGAAAGGCCCGCTGTGGGATACAGAGCTCAGCTGTGCGATAT 302
QY 144 GlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaProProLeuGly 163
DB 303 GGAGCCCCACCTCTTGATACGAGGCCGCCACCTGACAGATATGAGGCCGCCACCTTAGGA 362
QY 164 TyrGlyAlaProProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAlaProProLe 183
DB 363 TATGAGGCCCAACCTTTTGATATGAGGCCGCCACCTCTGAGATATGAGGCCGCCACCTTNT 422
QY 183 uGlyTyrGlyAlaProProAlaGlyAsnGluGlyProProAla 197
DB 423 CGGATATGAGGCCCAACCTCTGAGATATGAGGCCGCCCTTCG 465
RESULT 3
ID ABL96688 standard; cDNA; 467 BP.
XX ABL96688;
AC ABL96688;
XX
DT 21-JUN-2002 (first entry)
XX
DE Human testicular antigen encoding cDNA SEQ ID NO: 356.
XX
KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic; gene; ss.
OS Homo sapiens.
XX
XX
PN WO200155317-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR
```

```
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232168.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
```

PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240950.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483232/52.
 XX
 PT Nucleic acids encoding 973 human testicular antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating testicular cancer -
 XX
 PS Claim 1; SEQ ID NO 356; 766pp; English.
 XX
 CC The present invention provides the protein and coding sequences of 973

CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer,
 CC especially testicular cancers. The present sequence is a cDNA of the
 CC invention.
 XX
 SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
 Alignment Scores:
 Pred. No.: 3,68e-35 Length: 467
 Score: 826.00 Matches: 149
 Percent Similarity: 96.77% Conservative: 1
 Best Local Similarity: 96.13% Mismatches: 4
 Query Match: 64,588 Indels: 1
 DB: 23 Gaps: 0
 US-09-864-291-12 (1-235) x ABL96688 (1-467)
 QY 44 ArgAaenGlyAaspAlailegluphealaglneMetVallyAlaAlaSerAlaValAla 63
 DB 3 AGAAATGAGAGTGCCTTAATGATTTGCCAGTTGATGTGAAGCTGCTGCTGCC 62
 QY 64 ArgGlyPheProLeuArgThrIleuAsnAspITppheserSerMetGlylleTyrValIle 83
 DB 63 CGAGATTCCTCACTTAAGACCTTAATGACTGATGCTGATGAGAAATTAATTAAT 122
 QY 84 ThrGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleValTyrGlyAla 103
 DB 123 ACTGGGAAAGGGAATATGTGCACTCCAGATGCTTGTTCATTAATTCATGAGCC 182
 QY 104 ProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAlaProProAlaGlyTyrGly 123
 DB 183 CCACCTGACAGATATGAGCCCTCCCTGATTCGAGCCCTCCAGATATGAG 242
 QY 124 AlaGlnProValGlyAsnGluGlyProProValGlyTyrArgAlaSerProValArgTyr 143
 DB 243 GCCCAACCCGTAGAAATATAGAGCCGCTGTGGATATGAGAGCCCTGCTGCAAT 302
 QY 144 GlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaProProLeuGly 163
 DB 303 GAGAGCCCACTCTTGAATACGAGGCCCACTGACAGATATGAGCCCACTCTAGAG 362
 QY 164 TyrGlyAlaPro-ProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAlaProProle 183
 DB 363 TATGAGCCCCCAACCTTGTGATATGAGACCCCACTCGATATGAGGCCCACTT 422
 QY 183 uGlyTyrGlyAlaProProAlaGlyAsnGluGlyProProAla 197
 DB 423 CGATATGAGGCCCACTGACAGAAATGAGCCCGCTTGG 465
 RESULT 4
 AAS20601
 ID AAS20601 standard; cDNA; 1413 BP.
 AC AAS20601;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE DNA encoding bovine perinuclear theca 32 (PT32).
 XX
 KW Testicular WW domain binding protein; hWWP; perinuclear theca 32;
 KW PT32; contraceptive; fertility; oocyte activation; vaccine;
 KW globzoospermy; spermiogenesis; spermatocoe; tyrosine kinase; c-Yes;
 KW Immunoc contraceptive; bovine; gene; ss.
 XX
 OS Bos sp.
 XX
 XX Key Location/Qualifiers
 FH primer_bind 30..50
 FT /note= "a
 FT /note= "primer binding site for cDNA isolation. The
 FT sequence differs from that of the forward primer

```

FT      CDS                               36..977      given in AAS20603"
FT      /tag= b
FT      /product= "PT32"
FT      /note= "Perinuclear theca 32"
FT      protein_bind
FT      /tag= C
FT      /note= "Primer binding site for cDNA isolation. The
FT      sequence differs from that of the reverse primer
FT      given in AAS20604"
XX      WO200190185-A2.
XX      PD
XX      29-NOV-2001.
XX      PF
XX      25-MAY-2001; 2001WO-CA00738.
XX      PR
XX      25-MAY-2000; 2000CA-2307128.
XX      PR
XX      25-MAY-2000; 2000US-206979P.
XX      PA
XX      (TOOH ) UNIV QUEBENS KINGSTON.
XX      PA
XX      (UYOR-) UNIV OREGON HEALTH SCI.
XX      PI
XX      Oko R, Surovsky P;
XX      DR
XX      WPI, 2002-097644/13.
XX      DR
XX      P-PSDB; AAU74604.
XX      PT
XX      Isolated perinuclear theca 32 polypeptide that interacts with activated
XX      PT
XX      tyrosine kinase c-Yes, for enhancing fertility, treating/diagnosing
XX      PT
XX      diminished fertility and abnormal spermiogenesis and for providing
XX      PT
XX      contraception -
XX      PS
XX      Claim 10; Fig 2A-B; 103pp; English.
XX      CC
XX      The invention describes an isolated perinuclear theca 32 (PT32)
XX      CC
XX      polypeptide (1) which interacts with tyrosine kinase c-Yes. (1) is
XX      CC
XX      useful for: enhancing fertility in a mammal; creating glabrouspermy, by
XX      CC
XX      expressing (1) in spermatozoa; inhibiting fertilisation, by introducing
XX      CC
XX      (1) or its antigenic fragment into a mammal to elicit an immune
XX      CC
XX      response; enhancing the ability of round spermatids to activate oocytes;
XX      CC
XX      treating or diagnosing diminished fertility and abnormal spermiogenesis;
XX      CC
XX      in providing contraception; identifying contraceptive and
XX      CC
XX      fertility-enhancing agents. The polynucleotide is useful for producing
XX      CC
XX      (1) by recombinant techniques, as vaccine, as diagnostic reagents, and
XX      CC
XX      for chromosome identification. An antibody against (1) is useful in
XX      CC
XX      immunological assays, in immunoprecipitative methods, to identify cells
XX      CC
XX      expressing (1), and to purify (1) by affinity chromatography. A
XX      CC
XX      transgenic animal is useful as an animal model for studying human
XX      CC
XX      fertility and reproductive biology, and for screening compounds to
XX      CC
XX      identify modulators of oocyte activation. The use of (1) prevents the
XX      CC
XX      entry of components which are detrimental to embryonic development into
XX      CC
XX      the oocyte during oocyte activation with crude sperm extract and avoids
XX      CC
XX      the propagation of viruses such as HIV (human immunodeficiency virus) and
XX      CC
XX      SIV (simian immunodeficiency virus) carried in the sperm. This sequence
XX      CC
XX      encodes the bovine perinuclear theca 32 (PT32), described in the method
XX      CC
XX      of the invention.
XX      SQ
XX      Sequence 1413 BP; 377 A; 363 C; 369 G; 304 T; 0 other;
XX
XX      Alignment Scores:
XX      Pred. No.: 3..22e-32 Length: 1413
XX      Score: 777.00 Matches: 150
XX      Percent Similarity: 68.57% Conservative: 18
XX      Best Local Similarity: 61.22% Mismatches: 63
XX      Query Match: 60.75% Indels: 14
XX      DB: 24 Gaps: 4
XX
XX      US-09-864-291-12 (1-235) x AAS20601 (1-1413)
XX
XX      QY 1 MetProPheAspLeuThrAsnLeuThrValGluGlnProValPheAlaIaAsnPh 20
XX      DB 258 ATGCCGTTTGAGCTGATGAGTACGACCACTTGAACCAACCAATTTCGCCCAACTAC 317

```

```

QY 21 ILeYegLYThrIIeGlnAlaIaProTYrGlyYTrpGluGluGlnAlaThrPheLys 40
DB 318 ATTAAGGACCATTCAGGCGCTCCAGTGTGCTGGGAAGGACAGCTTTTAAG 377
QY 41 LeuValPheArgAsnGlyAspAlaIleGluPheAlaGlnLeuMetValIysAlaIaSer 60
DB 378 TTATCCCTTCAGGAAGAAGAGGTGCATGGAATTTCCCACTGATGTAAGAGCGCTCT 437
QY 61 AlAValAlaArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIle 80
DB 438 GCTGCTGCCAGAGGAATTCACCTTGGAAGGTAAATTAATGAGTTTCACACTTCAGGACTG 497
QY 81 TYrValIleThrArgLYGluGly---AsnMetCysThrProGlnMetProCys----- 96
DB 498 TACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 557
QY 97 SerValIleValIleTYrGlyAlaProProAlaGly-----TYrGly 109
DB 558 CCAATTTGATGATCTATGACACCCCAACACAGATATACAGTCAACAGGAGGATATGGA 617
QY 110 AlaProProProGlyTYrGlyAlaProProAlaGlyTYrGlyAlaGlnProValGlyAsn 129
DB 618 ACTCCACCAAGAGATATGAGCCCAACAGGAGGATATGAGCCCACTATGAGATAT 677
QY 130 GluGlyProProValGlyTYrArgAlaSerProValArgTYrGlyAlaProProLeuGly 149
DB 678 GAGGCCCGCTGAGGATATGAGTCCCACTGGAGATATGAGTCCCACTGGAGGA 737
QY 150 TYrGlyAlaProProAlaGlyTYrGlyAlaProProLeuGlyTYrGlyAlaProProLeu 169
DB 738 TATGAGTCCCACTGAGGATATGAGCCCACTGGAGATATGAGTCCCACTGGAG 797
QY 170 GlyTYrGlyThrProProLeuGlyTYrGlyAlaProProLeuGlyTYrGlyAlaProPro 189
DB 798 GGAATATGATGCCCACTGGAGATATGAGCCCACTGGAGATATGAGCCCACTGGAG 857
QY 190 AlaGlyAsnGluGlyProProAlaGlyTYrArgAlaSerProAlaGlySerGlyAlaArg 209
DB 858 GCTGAAATTAAGAGCCCTACCCCTGCATATAGCTTCATGCTGGAATATACAGCTGCC 917
QY 210 ProGlnGlnSerThrAlaAlaGlnAlaProGluAsnGluAlaSerLeuProSerAlaSer 229
DB 918 TCTCAACATCTATGACAGCTCAG-----CAGAGACTTCTTCCCACTACCTCA 968
QY 230 SerSerGlnValHis 234
DB 969 TCTTCTTAGTTCAT 983
XX
XX      RESULT 5
XX      ID ABRK84798 standard; cDNA; 220895 BP.
XX      XX
XX      ABRK84798;
XX      DT 14-AUG-2002 (first entry)
XX      XX
XX      Human cDNA differentially expressed in granulocytic cells #3369.
XX      XX
XX      Human; ss; granulocytic cell; DNA chip; bacterial infection;
XX      XX      viral infection; parasitic infection; protozoal infection;
XX      XX      fungal infection; sterile inflammatory disease; poriasis;
XX      XX      rheumatoid arthritis; glomerulonephritis; asthma; chromobis;
XX      XX      cardiac reperfusion injury; renal reperfusion injury; ARDS;
XX      XX      adult respiratory distress syndrome; inflammatory bowel disease;
XX      XX      Crohn's disease; ulcerative colitis; periodontal disease;
XX      XX      granulocyte activation; chronic inflammation; allergy.
XX      XX
XX      Homo sapiens.
XX      OS
XX      WO200228999-A2.
XX      PN
XX      11-APR-2002.
XX      PD

```

XX	03-OCT-2001; 2001WO-US30821.
PR	
XX	03-OCT-2000; 2000US-237189P.
FR	
XX	(GENE-) GENE LOGIC INC.
PA	
XX	
P1	Beazer-Barclay Y, Weisman SM, Yamaga S, Vockley J;
DR	WPI; 2002-435328/46.
XX	
PT	Detecting granulocyte activation by detecting differential expression
PR	of genes associated with granulocyte activation, which serves as
PT	diagnostic markers that is useful for monitoring disease states and
PT	drug toxicity -
XX	
PS	Claim 1; SEQ ID NO 1369; 114pp; English.
XX	
CC	The invention relates to detecting (M1) granulocyte (GC) activation
CC	(GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC	DNA chip analysis as given in the specification, and comparing
CC	the expression level to an expression level in an unactivated
CC	GC, where differential expression of Gs is indicative of GCA.
CC	Also included are modulating (M2) GA by contacting GC with an agent
CC	that alters the expression of at least one gene in Gs; (2) screening (M3)
CC	for an agent capable of modulating GCA or an inflammation (especially
CC	chronic) in a tissue, an allergic response in a subject, exposure of a
CC	subject to a pathogen or sterile inflammatory disease using the
CC	gene expression profile; (3) detecting (M4) an inflammation (especially
CC	chronic) in a tissue, an allergic response in a subject, exposure of a
CC	subject to a pathogen or sterile inflammatory disease, by detecting the
CC	level of expression in a sample of the tissue of gene(s) from Gs, where
CC	the level of expression of the gene is indicative of inflammation;
CC	(4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC	an allergic response in a subject, exposure of a subject to a pathogen
CC	or sterile inflammatory disease, by contacting a tissue having
CC	inflammation with an agent that modulates the expression of gene(s)
CC	from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC	modulating GA; M3 is useful for screening an agent capable of modulating
CC	GCA preferably in an inflammation in a tissue; M4 is useful for
CC	detecting an inflammation (especially chronic) in a tissue, an allergic
CC	response in a subject, exposure of a subject to a pathogen or sterile
CC	inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC	glomerulonephritis, asthma, thromboisis, cardiac reperfusion injury, renal
CC	reperfusion injury, ARDS, adult respiratory distress syndrome,
CC	inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC	periodontal disease; also bacterial infection, viral infection,
CC	parasitic infection, protozoal infection, fungal infection and M5 is
CC	useful for treating one of the above conditions. The present
CC	sequence represents a gene differentially expressed in granulocytes.
CC	Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 220895 BF; 53760 A; 54808 C; 55831 G; 56496 T; 0 other;
XX	
Alignment Scores:	
.Pred. No.:	9, 74e-30 Length: 220895
Score:	765.00 Matches: 138
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	59.81% Indels: 0
DB:	24 Gaps: 0
US-09-864-291-12 (1-235) * ABR84798 (1-220895)	
QY	98 ValIIEValTYrGIyAlAProFroAlagIYYrGIyAlAProBProFIYYrGIyAlA 117
Db	170005 GTTATGTCTATGGAGCCCACTGACAGATTAGAGGCCCACTCCCGATTACGGAGCC 170066
QY	118 ProPrOALIGLYTYrGIyAlAgInProVALIGLyAsngLUdlyProProVALIGLYTrArg 137

Df	170065	CCACCTGAGATATGAGACC	CACCGTAGGAATGAAAGGCCCGCTGTGGATTACAGA	170124
Oy	138	AAsSerProVAlAgtTgYtGjVAIAseProPLeuNgLYTtTgYtVAIAseProPLeuAgLYTt		157
Df	170125	GcCtCAcCtTGgcAGATATGAGCCCCCACCTCTTGATATGAGAGCCCACCTCGAGATAT		170184
Oy	158	GIYAIAseProPLeuNgLYTtTgYtVAIAseProPLeuNgLYTtTgYtThrProPLeuNgLY		177
Df	170185	GGAAGCCCAcCCTTRAGATATGAGCCCCCACCTCTTGATATGAAACCCACCTCTCGGA		170244
Oy	178	TtTgYtVAIAseProPLeuNgLYTtTgYtVAIAseProPLeuAgLYAAbngLYtProPLeuA		197
Df	170245	TATGAGAGCCCAcCCTCTCGATATGAGAGCCCCACCTCGAGAAATGAAAGCCCGCTGGG		170304
Oy	198	GLYTtTAATAIAseProPLeuAgLYSertGlyVAIAagProIngLuseThrAlaIagin		217
Df	170305	GGATTCACAAGcCCTCAcCCTCGATATGAGAGCCCAcCCTCGAATTTACACAGCCCGAG		170364
Oy	218	AlAProgiUAAbngLYAIAsePLeuProSeRAlaseRserSerGlyValHIsSer		235
Df	170365	GCTCTGAAAAAGAGGcCTTCTCTTCCCTCTGcCTCTCTTCTCAAGTCACTT		170418
RESULT 6				
ID	ABA67850	ABA67850 standard; DNA; 436 BP.		
XX	ABA67850;			
AC				
DT	01-FEB-2002	(first entry)		
XX				
DE	Human foetal liver single exon nucleic acid probe #16155.			
XX				
KW	Human; foetal liver; gene expression; single exon nucleic acid probe; ss.			
XX				
OS	Homo sapiens.			
XX				
PN	WO200157277-A2.			
PD				
FP	09-AUG-2001.			
XX				
Pf	30-JAN-2001; 2001WO-US00669.			
XX				
PR	04-FEB-2000; 2000US-0180312.			
PR	26-MAY-2000; 2000US-0207456.			
PR	30-JUN-2000; 2000US-0608408.			
PR	03-AUG-2000; 2000US-0632366.			
PR	21-SEP-2000; 2000US-0234687.			
PR	27-SEP-2000; 2000US-0236359.			
PR	04-OCT-2000; 2000GB-0024263.			
PA	(MOLE-) MOLECTUAR DYNAMICS INC.			
XX				
PI	Penn SG, Hanzel DK, Chen W, Rank DR;			
XX				
DR	WPI; 2001-483447/52.			
PT	Human genome-derived single exon nucleic acid probes useful for			
XX	analyzing gene expression in human fetal liver -			
XX				
PS	Claim 4; SEQ ID NO 16155; 639pp + sequence listing; English.			
XX				
CC	The invention relates to a single exon nucleic acid probe for			
CC	measuring human gene expression in a sample derived from human foetal			
CC	liver. The single exon nucleic acid probes may be used for predicting,			
CC	measuring and displaying gene expression in samples derived from human			
CC	fetal liver. The present sequence is a single exon nucleic acid			
CC	probe of the invention.			
CC	Note: The sequence data for this patent did not form part of the			
CC	printed specification, but was obtained in electronic format directly			
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.			
XX				
SO	Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;			

PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0255719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rozen CA, Barash SC, Ruben SM;
DR WPI; 2001-465570/50.
XX
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX
XX
XX Disclosure; SEQ ID NO 7570; 1297bp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,

CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
Alignment Scores:
Pred. No.: 6,71e-28 Length: 7099
Score: 705.00 Matches: 125
Percent Similarity: 100.00% Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 55.12% Indels: 0
DB: Gaps: 0
US-09-864-291-12 (1-235) x AAL04882 (1-7099)
QY 98 ValIleValIYrGlyAlaProProAlaGlyTYrGlyAlaProProProGlyTYrGlyAla 117
DB 6725 GTTATTGTCTATGAGGCCCACTGAGATATGAGACCCCACTCCGATACGAGACC 6784
QY 118 ProProAlaGlyTYrGlyAlaGlnProValGlyYrProProValGlyTYrArg 137
DB 6785 CCACTTCAGATATGAGCCCAACCCCTAGAAATGAAAGCCGCTGTGGATACGA 6844
QY 138 AlaSerProValArgTYrGlyAlaProProLeuGlyTYrGlyAlaProProAlaGlyTYr 157
DB 6845 GCCTCACCTGAGCATATGAGCCCACTTGTGATACGAGGCCCACTGCGAGATAT 6904
QY 158 GlyAlaProProLeuGlyTYrGlyAlaProProLeuGlyTYrGlyTYrProProLeuGly 177
DB 6905 GGAGCCCACTTATGAGATATGAGCCCACTTGTGATATGAGACCCCACTCTCGGA 6964
QY 178 TyrGlyAlaProProLeuGlyTYrGlyAlaProProAlaGlyYrProProAla 197
DB 6965 TATGAGCCCACTCTCGATATGAGCCCACTGCAAGAAATGAAAGCCGCTGCG 7024
QY 198 GlyTYrArgAlaSerProAlaGlySerGlyAlaArgProGlnGlnSerThrAlaAlaGln 217
DB 7025 GGATACAGAGCCTCACCTGCTGATCAGAGCCAGGCTCAGGAATCTACAGAGCCGAG 7084
QY 218 AlaProGlnGlnGlu 222
DB 7085 GCTCCTGAAAACGAG 7099
RESULT 11
ABL97776
ID ABL97776 standard; DNA; 7099 BP.
XX
XX ABL97776;
AC
XX
XX 21-JUN-2002 (first entry)
DT
XX
XX Human testicular antigen encoding DNA fragment SEQ ID NO: 2428.
DE
XX
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic; gene; db.
XX
XX Homo sapiens.
OS
XX
XX MO200155317-A2.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01329.
PF
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.

CC especially testicular cancers. The present sequence is a DNA encoding a
CC protein fragment of the invention.

XX
SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;

Alignment Scores:

Pred. No.:	6.71e-28	Length:	7099
Score:	705.00	Matches:	125
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	55.12%	Indels:	0
DB:	23	Gaps:	0

US-09-864-291-12 (1-235) x ABL77776 (1-7099)

```
OY 98 ValIleValIleTyrGlyAlaProProAlaGlyTyrGlyAlaProProGlyTyrGlyAla 117
DB 6725 GTTATGCTATGAGGCCCACTTCAGATATGAGGCCCACTCCCGATACCGAGCC 6784
OY 118 ProProAlaGlyTyrGlyAlaGlnProValGlyAangluGlyProProValGlyTyrArg 137
DB 6785 CCACTGCAAGATATGAGGCCCACTTCAGATATGAGGCCCACTCCCGATACCGAG 6844
OY 138 AlaSerProValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyr 157
DB 6845 GCTTCACTGTCGATATGAGGCCCACTTCGATATGAGGCCCACTTCGAGATAT 6904
OY 158 GlyAlaProProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyTyrProProLeuGly 177
DB 6905 GAGGCCCACTTCGATATGAGGCCCACTTCGATATGAGGCCCACTTCGAG 6964
OY 178 TyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyAangluGlyProProAla 197
DB 6965 TATGAGGCCCACTTCGATATGAGGCCCACTTCGAGATATGAGGCCCACTTCG 7024
OY 198 GlyTyrAlaSerProAlaGlySerGlyAlaArgProGlnGlySerThrAlaAlaGln 217
DB 7025 GGATACAGAGCTCACTGCTGATCAGAGGCCCACTTCGATACAGAGGCCCA 7084
OY 218 AlaProGluAanglu 222
DB 7085 GCTCCTGAAAACGAG 7099
```

RESULT 12

ID AAS70582 standard; cDNA; 894 BP.

AC AAS70582;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #6386.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSB-) HYSBQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG06395.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 1; SEQ ID NO 6386; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 894 BP; 204 A; 221 C; 213 G; 256 T; 0 other;

Alignment Scores:

Pred. No.:	6.95e-10	Length:	894
Score:	342.00	Matches:	69
Percent Similarity:	61.74%	Conservative:	2
Best Local Similarity:	60.00%	Mismatches:	12
Query Match:	26.74%	Indels:	32
DB:	23	Gaps:	2

US-09-864-291-12 (1-235) x AAS70582 (1-894)

```
OY 130 GluGlyProProValGlyTyrArgAlaSerProValArgTyrGlyAlaProProLeuGly 149
DB 168 GAGGCCGCCCTGCGATATGAGGCCCACTTCGATATGAGGCCCACTTCG 206
OY 150 TyrGlyAlaProProAlaGlyTyrGlyAlaProProLeuGlyTyrGlyAlaProProLeu 169
DB 207 -----TATGAGGCCCACTTC 224
OY 170 GlyTyrGlyTyrProProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyAlaProPro 189
DB 225 GGATATGAGAACCCCACTTCGATATGAGGCCCACTTCGATATGAGGCCCACT 284
OY 190 AlaGlyAangluGlyProProAlaGlyTyrArgAlaSerProAlaGlySerGlyAlaArg 209
DB 285 GCAGGAATATGAGGCCCGCTGCGATATGAGGCCCACTTCGATATGAGGCCGAG 344
OY 210 ProGlnGlySer-----ThrAlaAlaGlnAla 218
DB 345 CTTGAGGAATTCAGAGGAATATGAGGCCCACTTCGATATGAGGCCCACTATGAGG 404
OY 219 ProGluAangluAlaSerLeuProSerAlaSerSerGlnVal 233
DB 405 CCCCAAGAGGAGTCAAGCCCTGCTGAGAGCTCCCAAGTC 449
```

RESULT 13

ID AAS70430 standard; cDNA; 1885 BP.

AC AAS70430;

DT 27-NOV-2002 (first entry)

XX Human bone remodelling gene #87.
 DR Bone remodelling; osteoporosis; human; gene; ss.
 XX Homo sapiens.
 OS
 XX US6426186-B1.
 PN
 XX 30-JUL-2002.
 PD
 XX 18-JAN-2000; 2000US-0484970.
 PF
 XX 18-JAN-2000; 2000US-0484970.
 PR
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Jones KA, Volkmuth W, Walker MG;
 PI
 XX WPI; 2002-673014/72.
 DR
 XX A combination of polynucleotides which are co-expressed with genes
 PT known to be involved in bone remodeling and osteoporosis are useful in
 PT an array for the diagnosis of bone remodeling and osteoporosis
 PT associated disorders -
 XX
 XX Claim 1; Column 247-250; 206pp; English.
 XX
 XX The invention relates to a combination comprising a number of
 CC substantially purified and isolated polynucleotides which are
 CC co-expressed with genes known to be involved in bone remodeling and
 CC osteoporosis. The invention is used to diagnose disorders associated
 CC with bone remodeling or osteoporosis. ABS70344-ABS70512 represent
 CC human bone remodelling genes of the invention.
 XX
 SQ Sequence 1885 BP; 384 A; 610 C; 476 G; 415 T; 0 other;
 Alignment Scores:
 Pred. No.: 1,91e-05 Length: 1885
 Score: 261.50 Matches: 76
 Percent Similarity: 41.70% Conservative: 22
 Best Local Similarity: 32.34% Mismatches: 74
 Query Match: 20.45% Indels: 63
 DB: 24 Gaps: 9
 US-09-864-291-12 (1-235) x ABS70430 (1-1885)
 QY 1 MetProPheAAspleuMetThraenLeuThrValGluGlnProValPheAlaIaenphe 20
 DB 274 ATGCCATTATTATCTCATGAAGAAGCTGAGATCAAGACAGCCCTATTGTCGAACATAC 333
 QY 21 IleValGlyThrIleGlnAlaIaProValGlyGlyThrGluGlnAlaThrPheVal 40
 DB 334 ATCAAGGAAACAGTGAAGCGGAGCGGAGCTGCGGAGGAGCTGCTCTCTCAAG 393
 QY 41 LeuValPheAArgsnGlyAspAlaIleGluPheAlaGluLeuMetValIysAlaIaSer 60
 DB 394 TTGACTTTTCAAGCGGAGCGGCGCCATTGAGTTCGACAGGAGATGCTCCAGTGGCATCT 453
 QY 61 AlaValAlaArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIle 80
 DB 454 CAAGCTCCAGAGGT----- 468
 QY 81 TyrValIleThrGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
 DB 469 -----GAAGTCCCGCAGTGGAGCCCTATGAC 492
 QY 101 TyrGlyAlaProProAlaGlyTyrGlyAlaProProGlyTyrGlyAlaProProAla 120
 DB 493 TACTCTTACATGCGCGCGG-----GCCATATCTATCCCGG--- 531
 QY 121 GlyTyrGlyAlaGlnProValGlyAsnGluGlyProProValGlyTyrArgAlaSerPro 140

DB 532 -----CCAGTCGCCAAT-----GGAAATGACCCCTGCCCT 561
 QY 141 ValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaPro 160
 DB 562 CCTGGCTACCCCTATCCACCG-----CCCCACCTGAGTTCTATCCAGACCC 609
 QY 161 ProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAla 180
 DB 610 CCATGATGAGACGGGGCC-----ATGGAGATACGTGACGCCCA----- 648
 QY 181 ProProLeuGlyTyrGlyAlaProProAlaGlyAsnGluGlyProProAlaGlyTyrArg 200
 DB 649 CCACCG-----CCCTACCCCTGGCGCCATGAAACCTCCGGTTCAGCGGCC 693
 QY 201 AlaserProAlaGlySerGlyAlaArgProGlnGluSerThrAlaAlaGlnAlaProGlu 220
 DB 694 CGATGTCCTTCACCTCTGACGCGAAGCCAA-----GGCGCGAAGACAGCCGC 744
 QY 221 AsnGluAlaSerLeuProSerAlaSerSerGlnValHisSer 235
 DB 745 CAGCGCCTATTACAAACCAAGGCAATCTCACAGCTCTACATGCC 789
 RESULT 14
 AAF22363
 ID AAF22363 standard; cDNA; 1915 BP.
 XX
 AC AAF22363;
 AC
 DT 26-MAR-2001 (first entry)
 DT
 XX Human secreted protein gene 48 SEQ ID NO:58.
 DR
 XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; vulnary; gene therapy; neoplasm;
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischemia; angiogenesis; nervous system disorder; infection;
 KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;
 KW skin aging; food additive; preservative; ss.
 OS Homo sapiens.
 PN WO200061748-A1.
 PN
 XX 19-OCT-2000.
 PD
 XX 06-APR-2000; 2000WO-US08982.
 PF
 XX 09-APR-1999; 99US-0128696.
 PR 14-JAN-2000; 2000US-0176069.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 PI Rosen CA, Ruben SM, Komatsoulis G;
 XX WPI; 2000-638566/61.
 DR P-PSDB; AAB63096.
 DR
 XX New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosis, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 PS Claim 1; Page 429-430; 480pp; English.
 XX
 XX AAF22316 to AAF22363 encode the human secreted proteins given in AAB63049
 CC to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins
 CC and polypeptides homologous to them. Human secreted proteins have
 CC activities based on the tissues and cells the genes are expressed in.
 CC Examples of activities include: immunosuppressive; antiarthritic;
 CC antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;
 CC cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;

CC fungicide; ophthalmological; and vulnerary. The polynucleotides and
 CC proteins can be used to prevent, treat or ameliorate a medical condition
 CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
 CC sheep. They are also used in diagnosing a pathological condition or
 CC susceptibility to a pathological condition. Disorders which are diagnosed
 CC or treated include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
 CC ocular disorders e.g. corneal infection. The polypeptides can also be
 CC used to aid wound healing and epithelial cell proliferation, to prevent
 CC skin aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis. The polypeptides can also be used as a food additive or
 CC preservative to increase or decrease storage capabilities, fat content,
 CC lipid, protein, carbohydrate, vitamins, minerals, cofactors and other
 CC nutritional components. AAF22307 to AAF22315 and AAB6048 represent
 CC sequences used in the exemplification of the present invention.

XX Sequence 1915 BP, 401 A, 613 C, 481 G, 413 T, 7 other;

Alignment Scores:

Pred. No.:	2,94e-05	Length:	1915
Score:	258.00	Matches:	77
Percent Similarity:	42.13%	Conservative:	22
Best Local Similarity:	32.77%	Mismatches:	73
Query Match:	20.17%	Indels:	64
DB:	21	Gaps:	8

US-09-864-291-12 (1-235) x AAF22363 (1-1915)

```

OY 1 MetProPheAPLeuMetThrAnuLeuThrValGluGlnProValPheAlaAlaSerPhe 20
DB 272 ATGCATTATTTATCTCATGAAGAAGCTGTGAGATCATGACAGCCCGTATTGGTCAAACTAC 331
OY 21 ILeuValPheArgAsnGlyValAspAlaIleGluPheAlaGlnLeuMetValIleValAlaSer 40
DB 332 ATCAAGGAGAACGTAAGCGGAGACCGGAGCGGTGCTGGAGAAAGCTCTGCTTCTTCAACAG 391
OY 41 LeuValPheArgAsnGlyValAspAlaIleGluPheAlaGlnLeuMetValIleValAlaSer 60
DB 332 TTGACTTTACACGCGAGGGGGCCCATTAAGTTGCGACAGCGAGTCTCCAGGTGCATCT 451
OY 61 AlaValAlaArgGlyPheProLeuArgThrLeuAsnAspTyrPheSerSerMetGlyIle 80
DB 452 CAAGCTCCAGAGGT----- 466
OY 81 TyrValIleThrGlyGluGlyValAsnMetCysThrProGlnMetProCysSerValIleVal 100
DB 467 ----- 490
OY 101 TyrGlyAlaProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAlaProProAla 120
DB 491 TACTCTTACATGCCAGCGGSSCYTATGTCTATCCCGG----- 529
OY 121 GlyTyrGlyAlaGlnProValGlyValAsnGluGlyProProValGlyTyrArgAlaSerPro 140
DB 530 ----- 559
OY 141 ValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaPro 160
DB 560 CCTGGCTACCCCTATCCACCG----- 607
OY 161 ProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAla 180
DB 608 CCATGATGAGACGGGGCC----- 646
OY 181 ProProLeuGlyTyrGlyAlaProProAlaGlyValAsnGluGlyProProAlaGlyTyrArg 200
DB 647 CCACCG----- 691
OY 201 AlaSerProAlaGlySerGlyAlaArgProGlnGluSerThrAlaGlnAlaProGlu 220

```

```

DB 692 GATGCCCTCTCACTCTGACGCC-----GAGCGAAGCGGAGAGACGC 741
OY 221 AsnGluAlaSerLeuProSerAlaSerSerSerGlnValHisSer 235
DB 742 CAGCGCCTATTATCAACCCAGGCAATCTCTCAACAGCTTACATGCC 786

```

RESULT 15

ID AB273614 standard; cDNA; 1915 BP.

AC AB273614;

DT 12-MAY-2003 (first entry)

DS Secreted protein-encoding gene 334 cDNA clone HTLEMI5, SEQ ID NO:344.

XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antianaemic; vulnerary; gene; ss.

XX Homo sapiens.

OS WO200277013-A2.

PN 03-OCT-2002.

PD 26-MAR-2002; 2002WO-US09370.

PF 27-MAR-2001; 2001US-278650P.

PR 12-SEP-2001; 2001US-0950082.

PR 12-SEP-2001; 2001US-0950083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX MPI; 2003-040578/03.

XX P-PSDB, ABR01280.

XX New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorder, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia -

XX Claim 21; Page 1332; 2474pp; English.

XX AB273281-AB273697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC AB273698-AB274687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signaling, cellular differentiation, cell
 CC migration, prohormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein-encoding cDNA clone of the invention.

Sequence 1915 BP, 401 A, 613 C, 481 G, 413 T, 7 other;

Alignment Scores:

Pred. No.:	2,94e-05	Length:	1915
Score:	258.00	Matches:	77
Percent Similarity:	42.13%	Conservative:	22
Best Local Similarity:	32.77%	Mismatches:	73
Query Match:	20.17%	Indels:	64
DB:	25	Gaps:	8

US-09-864-291-12 (1-235) x AB273614 (1-1915)

```

QY      1 MetProPheAspLeuMetThrAsnLeuThrValGluGlnProValPheAlaAsnPhe 20
DB      272 ATGCCATTATCTCATGAAGACGTGAGATCAAGACAGCCCGTATTGGTGCAAACTAC 331
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      21 IleValGlyThrIleGlnAlaAlaProTyrGlyTyrGluGlyGlnAlaThrPheLeu 40
DB      332 ATCAAGGGAACGCTGAAGCGGAGCGGAGGTGGCTGGAGAGGCTCTGCTTCTACAG 391
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      41 LeuValPheArgAsnGlyAspAlaIleGluPheAlaGlnLeuMetValIlybAlaAser 60
DB      392 TTACCTTTCACGGCAGGGGGCGCCATTGAGTTGACAGCGGATGCTCCAGGTGGCATCT 451
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      61 AlaValAlaArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIle 80
DB      452 CAAGCTCCAGAGGT-----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      81 TyrValIleThrGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
DB      467 -----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      101 TyrGlyAlaProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAlaProProAla 120
DB      491 TACTCTTACATGCCCAAGGGGSCCTATGCTATCCCCG-----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      121 GlyTyrGlyAlaGlnProValGlyAsnGluGlyProProValGlyTyrArgAlaSerPro 140
DB      530 -----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      141 ValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaPro 160
DB      560 CCTGGCTACCCCTATCCACCG-----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      161 ProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAla 180
DB      608 CCCATATGACCGGGGCC-----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      181 ProProLeuGlyTyrGlyAlaProProAlaGlyAsnGluGlyProProAlaGlyTyrArg 200
DB      647 CCACCG-----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      201 AlaSerProAlaGlySerGlyAlaArgProGlnGluSerThrAlaAlaGlnAlaProGlu 220
DB      692 GATGTCCCTCCACTCCTGCAAGCC-----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      221 AsnGluAlaSerLeuProSerAlaSerSerGlnValHisSer 235
DB      742 CAGCGCTATTACCAACCCAGGCAATCTCTCAACAGTCTTACATGCC 786
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: December 16, 2003, 18:09:00
 Job time : 261.834 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 17:36:35 ; Search time 57.0347 Seconds

(without alignment)
1818.631 Million cell updates/sec

Title: US-09-864-291-12

Perfect score: 1279

Sequence: 1 MPDLMNTLVTEQVFAANF.....AQAPENBASLPSSSSQVHS 235

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlp
-O=/cgn2_1/USPRO_epool_P/US09864291/runat_15122003_160859_23332/app_query.fasta_1.846
-DB=Issued Patents NA -OFMT=fastap -SUFFIX=rml -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -STAFF=1 -END=-1 -MATRIX=bloum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09864291 @cgn 1 1 115 @runat_15122003_160859_23332 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FAPOP=6
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	261.5	20.4	1885	4	US-09-484-970B-87 Sequence 87, Appl
2	206.5	16.1	756	4	US-08-642-255-50 Sequence 50, Appl
3	196	15.3	654	3	US-08-998-416-1144 Sequence 1144, Ap
4	192	15.0	4403765	3	US-09-103-840A-2 Sequence 2, Appl
5	192	15.0	4411529	3	US-09-103-840A-1 Sequence 1, Appl
6	188.5	14.7	432	1	US-08-642-255-48 Sequence 48, Appl
7	186	14.5	2144	3	US-08-834-306-15 Sequence 15, Appl
8	186	14.5	2144	4	US-08-993-674A-15 Sequence 15, Appl
9	186	14.5	2144	4	US-09-256-976-15 Sequence 15, Appl
10	182.5	14.3	744	4	US-09-183-861-60 Sequence 60, Appl
11	182.5	14.3	744	4	US-09-022-765-60 Sequence 60, Appl
12	182.5	14.3	744	4	US-09-551-974A-60 Sequence 60, Appl

13	181	14.2	1235	2	US-08-557-309B-53 Sequence 53, Appl
14	166	13.0	2830	2	US-09-010-928B-1 Sequence 1, Appl
15	164.5	12.9	2824	2	US-09-010-928B-3 Sequence 3, Appl
16	162.5	12.7	2344	4	US-09-347-878-31 Sequence 31, Appl
17	162.5	12.7	3171	4	US-09-169-768-19 Sequence 19, Appl
18	162	12.7	810	1	US-08-642-255-60 Sequence 60, Appl
19	159.5	12.5	774	3	US-08-956-307B-12 Sequence 12, Appl
20	159.5	12.5	778	3	US-08-956-307B-11 Sequence 11, Appl
21	159.5	12.5	2428	1	US-08-445-050-1 Sequence 1, Appl
22	159.5	12.5	2428	1	US-08-204-691-1 Sequence 1, Appl
23	159.5	12.5	2428	4	US-09-355-295B-8 Sequence 2, Appl
24	159.5	12.5	2487	3	US-08-370-223-12 Sequence 12, Appl
25	159.5	12.5	2493	1	US-07-977-434-5 Sequence 5, Appl
26	159.5	12.5	2493	1	US-08-458-819-5 Sequence 5, Appl
27	159.5	12.5	3018	5	PCT-US91-07035-5 Sequence 5, Appl
28	159.5	12.5	3018	1	US-08-347-718B-3 Sequence 3, Appl
29	159.5	12.5	3018	1	US-08-482-262-3 Sequence 3, Appl
30	159.5	12.5	3018	6	US-08-068-945A-1 Sequence 1, Appl
31	159.5	12.5	11531	1	US-08-068-945A-1 Sequence 1, Appl
32	159.5	12.5	11531	1	US-08-442-806-1 Sequence 1, Appl
33	159.5	12.5	11531	4	US-09-355-295B-1 Sequence 1, Appl
34	157.5	12.3	4167	4	US-09-169-768-12 Sequence 12, Appl
35	157	12.3	2681	3	US-08-928-213B-7 Sequence 7, Appl
36	156	12.2	3170	4	US-09-169-768-1 Sequence 1, Appl
37	156	12.2	3171	4	US-09-169-768-15 Sequence 15, Appl
38	156	12.2	3181	1	US-08-655-086-1 Sequence 1, Appl
39	156	12.2	3349	4	US-09-169-768-13 Sequence 13, Appl
40	156	12.2	3531	4	US-09-169-768-7 Sequence 7, Appl
41	156	12.2	3541	4	US-09-169-768-9 Sequence 9, Appl
42	155.5	12.2	2496	1	US-08-073-384C-2 Sequence 2, Appl
43	155.5	12.2	2496	1	US-08-254-159A-2 Sequence 2, Appl
44	155.5	12.2	2496	1	US-08-483-043-2 Sequence 2, Appl
45	155.5	12.2	2496	1	US-08-481-238-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-484-970B-87
; Sequence 87, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmutch, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 87
; LENGTH: 1885
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6426186 232968.10CB1
US-09-484-970B-87

Alignment Scores:

Pred. No.: 2.69e-12 Length: 1885
Score: 261.50 Matches: 76
Percent Similarity: 41.70% Conservative: 22
Best Local Similarity: 32.34% Mismatches: 74
Query Match: 20.45% Indels: 63
DB: Gaps: 9

US-09-864-291-12 (1-235) x US-09-484-970B-87 (1-1885)

Oy 1 MePProheapleuMetThrAsnLeuThVaGluGlnProValPheAlaAspPhe 20
Db 274 ATGCCATTATCTCATGAAAGACTGTGAGATCAAGCGCCGATTGTGCTCAACTAC 333

APPLICANT: FRASER, Claire M.
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICANT: VENTNER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: TUBERCULOSIS
CURRENT APPLICATION NUMBER: US/09/103,840A
FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO: 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 0.0154 Length: 4411529
Score: 192.00 Matches: 58
Percent Similarity: 46.56% Conservative: 3
Best Local Similarity: 44.27% Mismatches: 30
Query Match: 15.01% Indels: 40
DB: 3 Gaps: 12

US-09-864-291-12 (1-235) x US-09-103-840A-1 (1-4411529)

QY 95 ProCysSerValIleValTyrGlyAlaProProAlaGlyTyrGlyAlaProPro----- 112
DB 4130468 CCGGAGCATGCTG-----CCGCTGGCTACCGGTTGAACACACCGCC 4130512
QY 113 -----ProGlyTyrGlyAlaProProAlaGlyTyrGlyAlaGlnProValGlyAmlu 130
DB 4130513 GTGGCGCGGAGTATGCG-----CCGGCGGATATCCGCGCTACCC----- 4130554
QY 131 GlyProProValGlyTyrTyrGlyAlaSerProValArgTyrGlyAlaProProleuGlyTyr 150
DB 4130555 GGTACACACACCGGGTAC-----GGCCCGCGGTTATGTCGCCCGCC--AGCTAT 4130605
QY 151 GlyAlaProProAlaGlyTyrGlyAlaProProleuGlyTyrGlyAlaProProleuGly 170
DB 4130606 GGC-----CCCGCGCTGGCTATAGT---CCACCCCTCGGCTACCCCGCGCACCGCGGC 4130659
QY 171 TyrGlyTyrProProleuGlyTyrGlyAlaProProleuGlyTyrGlyAlaProProAla 190
DB 4130660 TCGGCGCCACCGCC--GCTATAGC---CCACCGCTCGGCTATAGCCACCGGTCGCC 4130713
QY 191 GlyAmluGlyProPro----- 196
DB 4130714 CCGGCGCGGTCACACGAGATATATCCGCTGCCGCGCTTGAAGCATATCTTC 4130773
QY 197 -----AlaGlyTyr---ArgAlaSerPro 203
DB 4130774 AACGGCGCGTGGCTACATCGCGCTAACCG 4130806

RESULT 6
US-08-642-255-48/c
Sequence 48, Application US/08642255
Patent No. 5773249
GENERAL INFORMATION:
APPLICANT: CAPPELLO, Joseph
APPLICANT: FERRARI, Franco A.
TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSES: FLHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICANT: VENTNER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: TUBERCULOSIS
CURRENT APPLICATION NUMBER: US/08/642,255
FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO: 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 3.04e-07 Length: 432
Score: 188.50 Matches: 70
Percent Similarity: 52.52% Conservative: 3
Best Local Similarity: 50.36% Mismatches: 35
Query Match: 14.74% Indels: 32
DB: 1 Gaps: 14

US-09-864-291-12 (1-235) x US-08-642-255-48 (1-432)

QY 82 ValIleThrGlyGlyGlyAmluMetCysThrProGlnMetProCysSerValIleValTyr 101
DB 400 GTCTCTGGGCTATGAGCACCTGCGGG--CTTCAGATGCC----- 360
QY 102 GlyAlaPro-----ProAlaGlyTyrGlyAlaPro-----ProProGlyTyrGlyAla 117
DB 359 GAGACACCGGTTGACGACCGGTCGACGAGGCACTGGCGGCTGAGAGTCCCGAGGCA 300
QY 118 Pro-----ProAlaGlyTyrGlyAlaGlnProValGlyAmluGlyProProValGly 135
DB 299 CCGGTTGACGACCGGTCGACGAGGCA---CTGGCGGCTGAGGTC---GGA 249
QY 136 TyrArgAlaSerProValArgTyrGlyAlaPro-----ProleuGlyTyrGlyAlaPro 153
DB 248 GACCCCGGTGACACCGCGGTGACGAGGCACTGGCGGCTGAGAGTCCCGAGGCACTC 189
QY 154 -----ProAlaGlyTyrGlyAlaPro-----ProleuGlyTyrGlyAlaPro----- 167
DB 188 GGTGACACGCGGTGACGAGGCACTGGCGGCTGAGAGTCCCGAGGCACTC 129
QY 168 ProleuGlyTyrGlyTyrPro-----ProleuGlyTyrGlyAlaPro-----Proleu 183
DB 128 CAGCGCGGTGACGAGGCACTGGCGGCTGAGAGTCCCGAGGCACTC 69
QY 184 GlyTyrGlyAlaPro-----ProAlaGlyAmluGlyPro-----ProAlaGly 198
DB 68 GGTGACGAGGCACTTTCGGGCTGCGGCTGAGGCACTGCGGCTTTCGAGGCTGCGG 12

RESULT 7
US-08-834-306-15
Sequence 15, Application US/08834306
Patent No. 6054135
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yaelir A.W.
TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSES: FLHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,306
FILING DATE: 15-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 2144 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-834-306-15

Alignment Scores:
Pred. No.: 3,54e-06 Length: 2144
Score: 186.00 Matches: 51
Percent Similarity: 40.91% Conservative: 3
Best Local Similarity: 38.64% Mismatches: 71
Query Match: 14.54% Indels: 7
DB: 3 Gaps: 1

US-09-864-291-12 (1-235) x US-08-834-306-15 (1-2144)

QY 84 ThrGlyGluGlyAanMetCyethrProGlnMetProCySerVal----- 98
DB 1211 ACCCGCGAAGGCTGCTGCTGCAACCGCGAAGGCGCTGCTCCACCGCGAAGACCGCTGC 1270
QY 99 ---11eValTyrgLy-AlaProProAlaGlyTyrgLyAlaProProGlyTyrgLyAl 117
DB 1271 TGCACCCCGGAAGGCTGCTGCACTGCCAAGGCTGCTCCACCGCGAAGGCTGCTGC 1330
QY 117 aProProAlaGlyTyrgLyAlaGlnProValGlyAanGluGlyProProValGlyTyra 137
DB 1331 TCCACCCCGGAAGACCGCTGCTCCACCGCGAAGACCGCTGCTCCACCGCGAAGGCTGC 1390
QY 137 gAlaSerProValaGlyTyrgLyAlaProProLeuGlyTyrgLyAlaProProAlaGlyTy 157
DB 1391 TGCTCAACCGCGAAGGCGCTGCTCCACCGCGAAGGCGCTGCTCCACCGCGAAGGC 1450
QY 157 rGlyAlaProProLeuGlyTyrgLyAlaProProLeuGlyTyrgLyThrProProLeuG 177
DB 1451 CGCTGCTGCAACCGCGAAGGCGCTGCTCCACCGCGAAGGCTGCTCCACCGCGAA 1510
QY 177 yTyrgLyAlaProProLeuGlyTyrgLyAlaProProAlaGlyAanGluGlyProProAl 197
DB 1511 GGCGCTGCTCCACCGCGAAGGCTGCTCCACCGCGAAGGCTGCTCCACCGCGC 1570
QY 197 aGlyTyraGAlaSerProAlaGlySerGlyAla 208
DB 1571 GAAGGCTGCTGCTGCTCCGTTGAAAGAGGCT 1604

RESULT 8
US-08-993-674A-15
; Sequence 15, Application US/08993674A

Patent No. 6228372
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 2144 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-993-674A-15

Alignment Scores:
Pred. No.: 3,54e-06 Length: 2144
Score: 186.00 Matches: 51
Percent Similarity: 40.91% Conservative: 3
Best Local Similarity: 38.64% Mismatches: 71
Query Match: 14.54% Indels: 7
DB: 3 Gaps: 1

US-09-864-291-12 (1-235) x US-08-993-674A-15 (1-2144)

QY 84 ThrGlyGluGlyAanMetCyethrProGlnMetProCySerVal----- 98
DB 1211 ACCCGCGAAGGCTGCTGCTGCAACCGCGAAGGCGCTGCTCCACCGCGAAGACCGCTGC 1270
QY 99 ---11eValTyrgLy-AlaProProAlaGlyTyrgLyAlaProProGlyTyrgLyAl 117
DB 1271 TGCACCCCGGAAGGCTGCTGCACTGCCAAGGCTGCTCCACCGCGAAGGCTGCTGC 1330
QY 117 aProProAlaGlyTyrgLyAlaGlnProValGlyAanGluGlyProProValGlyTyra 137
DB 1331 TCCACCCCGGAAGACCGCTGCTCCACCGCGAAGACCGCTGCTCCACCGCGAAGGCTGC 1390
QY 137 gAlaSerProValaGlyTyrgLyAlaProProLeuGlyTyrgLyAlaProProAlaGlyTy 157
DB 1391 TGCTCAACCGCGAAGGCGCTGCTCCACCGCGAAGGCGCTGCTCCACCGCGAAGGC 1450
QY 157 rGlyAlaProProLeuGlyTyrgLyAlaProProLeuGlyTyrgLyThrProProLeuG 177
DB 1451 CGCTGCTGCAACCGCGAAGGCGCTGCTCCACCGCGAAGGCTGCTCCACCGCGAA 1510
QY 177 yTyrgLyAlaProProLeuGlyTyrgLyAlaProProAlaGlyAanGluGlyProProAl 197

Db 1511 GGCGCTGCTCCACCGCGAAGGCTGCTCCACCGCGAAGGCTGCTCCACCGCG 1570
QY 197 AGLYTARGAlaserProAlaIyserylAla 208
Db 1571 GAAGGCTGCTGCTCCCGTGTGGAAAGAGGCT 1604

RESULT 9
US-09-256-976-15
Sequence 15, Application US/09256976
Patent No. 6419933
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
OF T. CRUZI INFECTION
FILE REFERENCE: 210121.422C3
CURRENT FILING DATE: 1999-02-24
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 2144
TYPE: DNA
ORGANISM: Trypanosoma cruzi
US-09-256-976-15

Alignment Scores:
Pred. No.: 3 54e-06 Length: 2144
Score: 186.00 Matches: 51
Percent Similarity: 40.91% Conservative: 3
Best Local Similarity: 38.64% Mismatches: 71
Query Match: 14.54% Indels: 7
DB: 4 Gaps: 1

US-09-864-291-12 (1-235) x US-09-256-976-15 (1-2144)

QY 84 ThrtGlyGluGlyAsnMetCysThrProGlnMeProCysSerVal----- 98
Db 1211 ACCCGGAGAGCTCTCTGTGACCCCGGAGGCGCGTCTCCACCGCGAAGCGCTGC 1270
QY 99 ---11eValTyrgly-AlaProProAlaIyTyrglyAlaProProAlaIyTyrglyAl 117
Db 1271 TGCACCGCGAAGGCTGCTGCACTGCCAAGGCTGCTCCACCGCGAAGGCTGCTGC 1330
QY 117 AProProAlaIyTyrglyAlaGlnProValGlyAsnGluGlyProProValGlyTyrAr 137
Db 1331 TCCACCGCGAAGCGCTGCTCCACCGCGAAGCGCTGCTCCACCGCGAAGGCTGC 1390
QY 137 gAlaserProValArTyrglyAlaProProAlaIyTyrglyAlaProProAlaIyTy 157
Db 1391 TGTCTCACCGCGAAGCGCTGCTCCACCGCGAAGGCGCTGCTCCACCGCGAAGGC 1450
QY 157 rGlyAlaProProAlaIyTyrglyAlaProProAlaIyTyrglyThrProProAla 177
Db 1451 CGGTGCTGACCGCGGAGGCGCGTGTGCCACCGCGAAGGCTGCTGCTCCACCGGAA 1510
QY 177 YTyrglyAlaProProAlaIyTyrglyAlaProProAlaIyAsnGluGlyProProAl 197
Db 1511 GCGCGTGTCTCCACCGCGAAGGCTGCTGCTCCACCGCGAAGGCTGCTGCTCCACCGC 1570
QY 197 aGlyTyrArGAlaserProAlaIySerylAla 208
Db 1571 GAAGGCTGCTGCTCTCTCCCTTGGAAAGAGGCT 1604

RESULT 10
US-09-183-861-60/c
Sequence 60, Application US/09183861
Patent No. 6365165
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, Davin C.
APPLICANT: Skelky, Yasir A.W.
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/183,861
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/022,765
FILING DATE: 12-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.420C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 744 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-183-861-60

Alignment Scores:
Pred. No.: 1 81e-06 Length: 744
Score: 182.50 Matches: 63
Percent Similarity: 39.18% Conservative: 4
Best Local Similarity: 36.84% Mismatches: 31
Query Match: 14.27% Indels: 74
DB: 4 Gaps: 11

US-09-864-291-12 (1-235) x US-09-183-861-60 (1-744)

QY 100 ValTyrglyAlaProProAlaIy-----TyrglyAlaProPro 112
Db 583 GTGTACGGCGCTCTCTTAGGGCCATCGCGCATTTTCTGTGTATGCGCTCTCT 524
QY 113 ProGly-----TyrglyAlaProProAlaIy----- 121
Db 523 TAGGGCCATCGCGCATTTTCTGTGTATGCGCGCTCTCTTAGGGCCATCGCGCAT 464
QY 122 -----TyrglyAlaGlnProValGlyAsnGluGlyProProValGlyTyrArGAla 138
Db 463 TTTTCTGTATGCGCGCT-----CTCTCTTAGGGCCATCGCGCT 425
QY 139 SerProValArTyrglyAlaProProAlaIy-----Tyrgly 151
Db 424 CATTTTCTGTATGCGCGCTCTCTTAGGGCCATCGCGCATTTTCTGTGTATGCGC 365
QY 152 AlaProProAlaIy-----TyrglyAlaProProAlaIy----- 163
Db 364 CGTCTCTTAGGGCCATCGCGCATTTTCTGTGTATGCGCGCTCTCTTAGGGCCAT 305
QY 164 -----TyrglyAlaProProAlaIy----- 170

DB 304 CGCGGTATTTTCTGTGTAGCGCGCTCTCTTAGGGCCATGCGCGTATTTTCTGTG 245
QY 171 TTTGTATTTTCTGTGTAGCGCGCTCTCTTAGGGCCATGCGCGTATTTTCTGTG 183
DB 244 TATGCGCGCTCTCTTAGGGCCATGCGCGTCTTTCTGTGTAGCGCGCTCTCTTAG 185
QY 184 GY-----TyrGlyAlaProProAlaGlyAsnGlu----- 193
DB 184 GCGCATGCGCGTATTTTCTGTGTATGCGCATCTCTCTTGGGGCA-GAGTCGTTGATC 126
QY 194 GlyProProAlaGlyTyrArgAlaSerProAla 204
DB 125 GCACCGCGCGCTGCTTGGGGGACACCGCGG 93

RESULT 11
US-09-022-765-60/c
Sequence 60, Application US/09022765
Patent No. 6375955
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, David C.
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,765
FILING DATE: 12-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.420C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6011
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 744 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-022-765-60

Alignment Scores:
Pred. No.: 1 81e-06 Length: 744
Score: 182.50 Matches: 63
Percent Similarity: 39.18% Conservative: 4
Best Local Similarity: 36.84% Mismatches: 31
Query Match: 14.27% Indels: 74
DB: 4 Gaps: 11

US-09-864-291-12 (1-235) x US-09-022-765-60 (1-744)

QY 100 ValTyrGlyAlaProProAlaGly-----TyrGlyAlaProPro 112
DB 583 GTGTAGCGCGCTCTCTTAGGGCCATGCGCGTATTTTCTGTGTATGCGCGCTCTCT 524
QY 113 ProGly-----TyrGlyAlaProProAlaGly----- 121

DB 523 TAGGGCCATGCGCGTATTTTCTGTGTAGCGCGCTCTCTTAGGGCCATGCGCGTAT 464
QY 122 -----TyrGlyAlaGlnProValGlyAsnGluGlyProProValGlyTyrArgAla 138
DB 463 TTTTCTGTGTAGCGCGCT-----CTCTTAGGGCCATGCGCGT 425
QY 139 SerProValAlaGlyTyrGlyAlaProProAlaGly-----TyrGly 151
DB 424 CATTTTCTGTGTAGCGCGCTCTCTTAGGGCCATGCGCGTATTTTCTGTGTAGCGC 365
QY 152 AlaProProAlaGly-----TyrGlyAlaProProAlaGly--- 163
DB 364 GTCCTCTTAGGGGACATGCGCGTATTTTCTGTGTATGCGCGCTCTCTTAGGGCCAT 305
QY 164 -----TyrGlyAlaProProAlaGly----- 170
DB 304 CGCGGTATTTTCTGTGTAGCGCGCTCTCTTAGGGCCATGCGCGTATTTTCTGTG 245
QY 171 TTTGTATTTTCTGTGTAGCGCGCTCTCTTAGGGCCATGCGCGTATTTTCTGTG 183
DB 244 TATGCGCGCTCTCTTAGGGCCATGCGCGTCTTTCTGTGTAGCGCGCTCTCTTAG 185
QY 184 GY-----TyrGlyAlaProProAlaGlyAsnGlu----- 193
DB 184 GCGCATGCGCGTATTTTCTGTGTATGCGCATCTCTCTTGGGGCA-GAGTCGTTGATC 126
QY 194 GlyProProAlaGlyTyrArgAlaSerProAla 204
DB 125 GCACCGCGCGCTGCTTGGGGGACACCGCGG 93

RESULT 12
US-09-551-974A-60/c
Sequence 60, Application US/09551974A
Patent No. 6500437
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, David C.
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
FILE REFERENCE: 210121.420C5
CURRENT APPLICATION NUMBER: US/09/551,974A
CURRENT FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 101
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 60
LENGTH: 744
TYPE: DNA
ORGANISM: Leishmania chagasi
US-09-551-974A-60

Alignment Scores:
Pred. No.: 1 81e-06 Length: 744
Score: 182.50 Matches: 63
Percent Similarity: 39.18% Conservative: 4
Best Local Similarity: 36.84% Mismatches: 31
Query Match: 14.27% Indels: 74
DB: 4 Gaps: 11

US-09-864-291-12 (1-235) x US-09-551-974A-60 (1-744)

QY 100 ValTyrGlyAlaProProAlaGly-----TyrGlyAlaProPro 112
DB 583 GTGTAGCGCGCTCTCTTAGGGCCATGCGCGTATTTTCTGTGTATGCGCGCTCTCT 524
QY 113 ProGly-----TyrGlyAlaProProAlaGly----- 121
DB 523 TAGGGCCATGCGCGTATTTTCTGTGTAGCGCGCTCTCTCTTAGGGCCATGCGCGTAT 464
QY 122 -----TyrGlyAlaGlnProValGlyAsnGluGlyProProValGlyTyrArgAla 138

Db 463 TTTTCGTGTACGCGCT-----CCTCTTAGGCGCATCGCGT 425
 QY 139 SerProValArgTyrGlyAlaProProLeuGly-----TyrGly 151
 Db 424 CATTTTGTGTGTACGCGCTCTCTTAGGCGCATCGCGTATTTTCTGTGTACGCG 365
 QY 152 AlaProProAlaGly-----TyrGlyAlaProProLeuGly--- 163
 Db 364 CGTCTCTTAGGCGCATCGCGCATTTTCTGTGTATGCGGTCTCTTAGGCGCAT 305
 QY 164 -----TyrGlyAlaProProLeuGly----- 170
 Db 304 CGCGTATTTTCTGTGTACGCGCTCTCTTAGGCGCATCGCGTATTTTCTGTG 245
 QY 171 TyrGlyThrProProLeuGly-----TyrGlyAlaProProLeu 183
 Db 244 TAGGCGCTCTCTCTTAGGCGCATCGGTCTTTCTGTGTATGCGGTCTCTTAG 185
 QY 184 Gly-----TyrGlyAlaProProAlaGlyAenglyu----- 193
 Db 184 GGCATCGCGCTCATTTTCTGTGTATGCGCATCTCTCTTAGGCGCA-GAGTGTTCATC 126
 QY 194 GlyProProAlaGlyTyrArgAlaSerProAla 204
 Db 125 GCAACGCGCGCTCTCTTAGGCGCATCGCGCG 93

RESULT 13
 US-08-557-309B-53
 ; Sequence 53, Application US/08557309B
 ; Patent No. 5916572

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yaeli A.W.
 APPLICANT: Lodes, Michael J.

APPLICANT: Houghton, Raymond L.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T

NUMBER OF SEQUENCES: 69

CORRESPONDENCE ADDRESS:

ADDRESSEE: SERD and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/557,309B

FILING DATE: 14-NOV-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121,422

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 53:

SEQUENCE CHARACTERISTICS:

LENGTH: 1235 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-557-309B-53

Alignment Scores:
 Pred. No.: 4.48e-06 Length: 1235
 Score: 181.00 Matches: 51
 Percent Similarity: 41.04% Conservative: 4
 Best Local Similarity: 38.06% Mismatches: 70

Query Match: 14.15% Indels: 9
 DB: 2 Gaps: 2

US-09-864-291-12 (1-235) x US-08-557-309B-53 (1-1235)

QY 84 ThrGlyGlyGlyAlaSerMetCysThrProGlnMetProCysSerVal-IleValTyrGlyAla 103
 Db 301 ACCCGGAAAGGCTGTGTGTACACCGGAAAGCGGTGTCTTCAACCGCGAAGAGCTTCC 360
 QY 103 aProPro-----AlaGlyTyrGlyAlaProProProGlyTyrGlyAla 117
 Db 361 TGCACCGCGAAGGCTGTGTGTACACCGGAAAGCGGTGTCTTCAACCGCGAAGAGCTTCC 420
 QY 117 aProProAlaGlyTyrGlyAlaGlnProValGlyAenglyuGlyProProValGlyTyr 137
 Db 421 TCCACCGCGAAGAGCGGTGTGTGTACACCGGAAAGCGGTGTCTTCAACCGCGAAGAGCTTCC 480
 QY 137 GalSerProValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyr 157
 Db 481 TGTCTTCAACCGCGAAGGCGGTGTGTGTACACCGGAAAGCGGTGTCTTCAACCGCGAAGG 540
 QY 157 rGlyAlaProProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyThrProProLeuG 177
 Db 541 CGCTGTGTGACCGCGAAGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
 QY 177 yTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyAenglyuGlyProProAl 197
 Db 601 GCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 660
 QY 197 aGlyTyrArgAlaSerProAlaGlySerGlyAlaArgPro 210
 Db 661 G-----GAGGCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 694

RESULT 14
 US-09-010-928B-1
 ; Sequence 1, Application US/09010928B
 ; Patent No. 5994099

GENERAL INFORMATION:

APPLICANT: Lewis, Randolph V

APPLICANT: Hayashi, Cheryl Y

TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA

TITLE OF INVENTION: CODING THEREFOR

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH

STREET: 8110 GATHEHOUSE RD. SUITE 500E

CITY: FALLS CHURCH

STATE: VIRGINIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 22042

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/010,928B

FILING DATE: 22-JAN-1998

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M

REGISTRATION NUMBER: 28977

REFERENCE/DOCKET NUMBER: 1447-109P

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2830 base pairs

TYPE: nucleic acid

STRANDEDNESS: not relevant

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: -

```

/ LOCATION: 1..2830
/ OTHER INFORMATION: /note="Flagelliform DNA sequence
/ OTHER INFORMATION: taken from the 5' region. The putative start codon is at
/ OTHER INFORMATION: position 219"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 219..2830
US-09-010-928B-1

Alignment Scores:
Pred. No.: 0.0002 Length: 2830
Score: 166.00 Matches: 57
Percent Similarity: 44.96% Conservative: 1
Best Local Similarity: 44.19% Mismatches: 41
Query Match: 12.98% Indels: 30
DB: 2 Gaps: 10

US-09-864-291-12 (1-235) x US-09-010-928B-1 (1-2830)
/
/
/ 95 ProCySerValIleValTyrGlyAlaProProAlaGlyTyrGlyAlaProProGly 114
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 990 CCGAGAGGTGTAGAACTGTGTAGACCGACCGAGCGATTACGGA-----CCTGGT 1040
/
/ 115 TyrGlyAlaProProAlaGlyTyrGlyAlaGlnProValGlyAsnGlnGlyProProVal 134
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1041 ---GGAGCCGACCGAGAGGTTATGACCTGTGTGAGCCGGA-----CCAGGA 1085
/
/ 135 GlyTyrAlaGlnSerProValArgTyrGlyAlaProProLeuGlyTyr-----Gly 151
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1086 GGTTACGAGCTGTGT-----GGAGCTGAGCGAGAGGTTACGACCTGTGTGA 1133
/
/ 152 AlaProProAlaGlyTyr-----GlyAlaProProLeuGlyTyr-----Gly 165
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1134 GCTGGGCTGTGAGGTTACGACCTGTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAG 1193
/
/ 166 AlaProProLeuGlyTyr-----GlyThrProProLeuGlyTyrGlyAlaProPro 182
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1194 GCTGACCTGTGAGGTTACGACCTGTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAG 1253
/
/ 183 LeuGlyTyrGlyAlaProProAlaGlyAsnGlnGlyPro-----ProAlaGly 198
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1254 ACTGACCTGTGAGGAGGTTGACCTGTGAGGAGCTGTGAGCGAGAGATATGACCTGTGTGT 1313
/
/ 199 TyrAlaGlnSerProAlaGlySerGly 207
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1314 -----GCTGACCTGTGTGTGTGTGA 1334
/
RESULT 15
US-09-010-928B-3
/ Sequence 3, Application US/09010928B
/ Patent No. 5994099
/ GENERAL INFORMATION:
/ APPLICANT: Lewis, Randolph V
/ APPLICANT: Hayashi, Cheryl Y
/ TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
/ TITLE OF INVENTION: CODING THEREFOR
/ NUMBER OF SEQUENCES: 29
/ CORRESPONDENCE ADDRESS:
/ ADDRESSES: BIRCH, STEWART, KOLASCH & BIRCH
/ STREET: 8110 CATERHOUSE RD. SUITE 500E
/ CITY: FALLS CHURCH
/ STATE: VIRGINIA
/ COUNTRY: UNITED STATES OF AMERICA
/ ZIP: 22042
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/010,928B
/ FILING DATE: 22-JAN-1998
/ CLASSIFICATION: 435

```

```

/ ATTORNEY/AGENT INFORMATION:
/ NAME: Murphy Jr., Gerald M
/ REGISTRATION NUMBER: 28977
/ REFERENCE/DOCKET NUMBER: 1447-109P
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2824 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: not relevant
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ HYPOTHEICAL: NO
/ FEATURE:
/ NAME/KEY: -
/ LOCATION: 1..2824
/ OTHER INFORMATION: /note="Flagelliform DNA sequence
/ OTHER INFORMATION: taken from 3' region. Stop codon begins at position 2722."
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..2724
US-09-010-928B-3

Alignment Scores:
Pred. No.: 0.000263 Length: 2824
Score: 164.50 Matches: 58
Percent Similarity: 42.67% Conservative: 6
Best Local Similarity: 38.67% Mismatches: 49
Query Match: 12.86% Indels: 37
DB: 2 Gaps: 11

US-09-864-291-12 (1-235) x US-09-010-928B-3 (1-2824)
/
/
/ 85 GlyGlnGlyAsnMetCysThrProGlnMetProCysSerValIleValTyrGlyAlaPro 104
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1273 GAGCCAGGTGTGTCTGTGCGGACCTTATGACACGAGAGGTGTAGACCTGTGTGAGTGA 1332
/
/ 105 ProAlaGlyTyr-----GlyAlaProProProGlyTyr-----GlyAlaPro 118
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1333 CCGGAGGTGTATGACCTGTGTGAGAGTGTGAGCTGTGAGGTTATGACCTGTGTGAGCTGGA 1392
/
/ 119 ProAlaGlyTyrGlyAlaGlnProValGlyAsnGlnGlyPro----- 132
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1393 CCGTGAAGTTACGACCT-----GCTGTTCTGTGTCCAGGTGTATGACCTGTGTGAGCTGGA 1443
/
/ 133 -----ProValGlyTyrAlaGlnSerProValArgTyrGlyAlaProProLeuGly 149
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1444 GATTCTGTGTCCAGGTGTATGACGACCCGGT-----GGTTCGAGCGACGAGAGA 1491
/
/ 150 Tyr-----GlyAlaProProAlaGlyTyr-----GlyAlaProProLeuGly 163
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1492 TACGACCTGTGCGGTTCTGTGACCTGTGTGATACGATCTGTGCGGCTGTGACCTGTGTGA 1551
/
/ 164 Tyr-----GlyAlaProProLeuGlyTyr-----GlyThrProProLeuGly 177
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1552 TACGACCTGTGCGGTTCTGTGACCTGTGTGATACGATCTGTGAGGTTCTGTGACCTGTGTGT 1611
/
/ 178 TyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyAsnGlnGlyProProAla 197
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1612 TATGACCTGTGTGTGTGTGTGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1668
/
/ 198 GlyTyrAlaGlnSerProAlaGlySerGly 207
/ DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
/ 1669 GGATAC-----GGACCTGTGTGTGTGTGTGA 1692

```

Search completed: December 17, 2003, 02:37:27
 DB time : 9986.04 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 17:58:55 ; Search time 229.854 Seconds

(without alignments)
3398.008 Million cell updates/sec

Title: US-09-864-291-12

Perfect score: 1279

Sequence: 1 MPPLMTNLTVEQVPANF.....AQAPENASLPSASSQVHS 235

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Pgapop 6.0 , Pgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 440344

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODL=frame+g2n.model -DEV=xip
-Q=/cgn2_1/USPRO_pool_P/US09864291/rnatc_15122003.160900.23360/app_query.fasta_1.846
-DB=Published Applications NA -QFMT=fastac -SUPFIX=mpb -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=10sum62
-TRANS=human40.cdd -LIST=45 -DOCMATCH=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALGN=15 -MODE=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09864291 @CGN 1.1.108 @rnatc_15122003.160900.23360
-NCPU=6 -ICPU=3 -NO MAP -LANG=ENGLISH -NEG SCORES=0 -WAIT -DISPATCH=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRBADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications NA:

1: /cgn2_6/ptodata/2/pubna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	826	64.6	467	11	US-09-764-891-1230	Sequence 1230, Ap
2	761	59.5	436	9	US-09-864-761-11561	Sequence 31561, A
3	705	55.1	7099	11	US-09-764-891-7570	Sequence 7570, Ap
4	256	20.0	511	13	US-10-029-386-24981	Sequence 24981, A
5	245	19.2	471	9	US-09-864-761-15029	Sequence 15029, A
6	226	17.7	511	13	US-10-029-386-24981	Sequence 24981, A
7	211	16.5	541	13	US-10-029-386-11265	Sequence 11265, A
8	202.5	15.8	1743	10	US-09-887-756-788	Sequence 788, Ap
9	197.5	15.4	2430	15	US-10-156-761-2531	Sequence 2531, Ap
10	197.5	15.4	9025608	15	US-10-156-761-1	Sequence 1, Appl
11	192.5	15.1	866	13	US-10-029-386-22839	Sequence 22839, A
12	192	15.0	321	11	US-09-764-891-1356	Sequence 1356, A
13	191	14.9	432	11	US-09-818-995-33536	Sequence 33536, A
14	185	14.5	780	13	US-10-029-386-20220	Sequence 20220, A
15	185	14.5	2016	10	US-09-338-842A-2004	Sequence 2004, Ap
16	184	14.4	507	11	US-09-770-861-263	Sequence 263, Ap
17	182.5	14.3	744	9	US-09-874-923-60	Sequence 60, Appl
18	182.5	14.3	744	10	US-09-991-496-60	Sequence 60, Appl
19	182.5	14.3	3760	10	US-09-887-576-458	Sequence 458, Appl
20	178.5	14.0	534	10	US-09-338-842A-968	Sequence 968, Ap
21	178.5	14.0	731	9	US-09-770-149-103	Sequence 103, Appl
22	178.5	14.0	1909	13	US-10-094-749-1166	Sequence 1166, Ap
23	176	13.8	959	13	US-10-029-386-22751	Sequence 22751, A
24	176	13.8	9025608	15	US-10-156-761-1	Sequence 1, Appl
25	175	13.7	1743	15	US-10-128-714-7526	Sequence 7526, Ap
26	173	13.5	2164	15	US-10-128-714-1526	Sequence 1526, Ap
27	173	13.5	2217	15	US-10-128-714-6526	Sequence 6526, Ap
28	173	13.5	4164	15	US-10-128-714-526	Sequence 526, Ap
29	173	13.5	4217	15	US-10-128-714-5526	Sequence 5526, Ap
30	172	13.4	1967	9	US-09-864-761-5097	Sequence 5097, Appl
31	170.5	13.3	2485	15	US-10-153-668-91	Sequence 91, Appl
32	169.5	13.3	685	13	US-10-029-386-20888	Sequence 20888, A
33	169	13.2	593	13	US-10-027-632-290832	Sequence 290832, A
34	169	13.2	593	14	US-10-027-632-290832	Sequence 290832, A
35	169	13.2	1674	15	US-10-156-761-3052	Sequence 3052, Ap
36	168.5	13.2	1040	13	US-10-029-386-26033	Sequence 26033, Ap
37	168	13.1	541	13	US-10-029-386-11265	Sequence 11265, A
38	166	13.0	631	13	US-10-029-386-20885	Sequence 20885, A
39	166	13.0	2108	10	US-09-962-832-225	Sequence 225, Appl
40	166	13.0	2709	15	US-10-156-761-3543	Sequence 3543, Ap
41	165.5	12.9	335	11	US-09-913-514-26	Sequence 26, Appl
42	165.5	12.9	648	13	US-09-864-761-21820	Sequence 21820, A
43	165.5	12.9	648	13	US-10-029-386-20230	Sequence 20230, A
44	165	12.9	409	10	US-09-960-352-6984	Sequence 6984, Ap
45	163	12.7	1470	14	US-10-029-180-75	Sequence 75, Appl

ALIGNMENTS

RESULT 1
US-09-764-891-1230
; Sequence 1230, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1230
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: SITE
; LOCATION: (421)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1230
Alignment Scores:

Pred. No.: 2.28e-71 Length: 467
 Score: 826.00 Matches: 149
 Percent Similarity: 96.77% Conservative: 1
 Best Local Similarity: 96.13% Mismatches: 4
 Query Match: 64.58% Indels: 1
 DB: 11 Gaps: 0

US-09-864-291-12 (1-235) x US-09-764-891-1230 (1-467)

Qy 44 ArgAsnGlyAspAlaIleGluPheAlaGlnLeuMetValIleValAlaValAla 63
 Db 3 AGAATGTGAGGTGTCATGTGATTTGCGCAGTTGTGTGAAAGCTGCTGCTGCGCC 62
 Qy 64 ArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIleTyrValIle 83
 Db 63 CGAGATTTCACCTTAAGACCTTAAGTGAAGCTGTTGAGCTTATGGAATTTATGTAAT 122
 Qy 84 ThrGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleValTyrGlyAla 103
 Db 123 ACTGGGAGAGGATATGTGACCTCCAGATGCTGTTGATTTATGTAAGGCC 182
 Qy 104 ProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAlaProProAlaGlyTyrGly 123
 Db 183 CCACTGAGAGATATGAGGCCCACTCCGATATAGAGGCCCACTGAGATATGGA 242
 Qy 124 AlaGlnProValGlyAsnGluGlyProProValGlyTyrTargAlaSerProValArgTyr 143
 Db 243 GCCCAACCCGTAGGAATGAAAGGCCCGCTGTGTGATACAGAGCTCACTGTGCGATAT 302
 Qy 144 GlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaProProLeuGly 163
 Db 303 GAGGCCCACTCTTGATGATGAGGCCCACTGAGATATGAGGCCCACTCTTGA 362
 Qy 164 TyrGlyAlaPro-ProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAlaProProLe 183
 Db 363 TAGGAGCCCAACCTTGTGATATGAGGCCCACTCTGATATATGAGGCCCACTTNT 422
 Qy 183 uGlyTyrGlyAlaProProAlaGlyAsnGluGlyProProAla 197
 Db 423 CGGATATGAGGCCCACTGAGGAATGAAAGGCCCGCTTGC 465

RESULT 2

US-09-864-761-31561
 ; Sequence 31561, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wenheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecm1ca-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263,6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 31561
 ; LENGTH: 436
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO Z99716.4
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
 ; OTHER INFORMATION: EST HUMAN HIT: AW169980.1, EVALU8 8.00e-11
 ; OTHER INFORMATION: SWISSPROT HIT: P18616, EVALU8 3.00e-13

US-09-864-761-31561

Alignment Scores:
 Pred. No.: 4.09e-65 Length: 436
 Score: 761.00 Matches: 137
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 59.50% Indels: 0
 DB: 9 Gaps: 0

US-09-864-291-12 (1-235) x US-09-864-761-31561 (1-436)

Qy 99 IleValTyrGlyAlaProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAlaPro 118
 Db 3 ATGTCTATGAGGCCCACTGAGATATGAGGCCCACTCCGATATGAGGCCCA 62
 Qy 119 ProAlaGlyTyrGlyAlaGlnProValGlyAsnGluGlyProProValGlyTyrTargAla 138
 Db 63 CTGCGAGATATGAGGCCCACTGAGGAATGAAAGGCCCGCTGTGATATGAGGCC 122
 Qy 139 SerProValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGly 158
 Db 123 TCACCTGTGCGATATGAGGCCCACTCTTGATATGAGGCCCACTGAGATATGGA 182
 Qy 159 AlaProProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyThrProProLeuGlyTyr 178
 Db 183 GCCCACTCTAGATATGAGGCCCACTCTTGATATGAGGCCCACTCTCGATAT 242
 Qy 179 GlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyAsnGluGlyProProAlaGly 198
 Db 243 GAGGCCCACTCTCGATATGAGGCCCACTGCGAGGAATGAAAGGCCCGCTGCGGGA 302
 Qy 199 TyrArgAlaSerProAlaGlySerGlyAlaArgProGlnGlySerThrAlaAlaGlnAla 218
 Db 303 TAGAGAGCTTCACTCTGATATGAGGCCCACTGAGATATGAGGCCCACTCTGAGCT 362
 Qy 219 ProGlyAsnGluAlaSerLeuProSerAlaSerSerSerGlnValIleSer 235
 Db 363 CTGGAATGAGGCTTCTCTTCCCTGCTGCTCTTCTTCAAGTCAATCT 413

RESULT 3

```

US-09-764-891-7570
; Sequence 7570, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7570
; LENGTH: 7099
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7570

Alignment Scores:
Pred. No.: 2.03e-58 Length: 7099
Score: 705.00 Matches: 125
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 55.12% Indels: 0
DB: 11 Gaps: 0

US-09-864-291-12 (1-235) x US-09-764-891-7570 (1-7099)
QY 98 ValIleValIYrGlyAlaProProAlaGlyYrGlyAlaProProGlyYrGlyAla 117
Db 6725 GTTATTGTCTATTGAGGCCCACTGTCAGATATGAGCCCACTCCGGAATGAGAGCC 6784
QY 118 ProProAlaGlyYrGlyAlaGlnProValGlyAsnGlyYrProProValGlyYrArg 137
Db 6785 CCACCTGAGGATATGAGCCCACTGAGAAATGAGGCCCTGTCGATACAGA 6844
QY 138 AlAserProValArgYrGlyAlaProProLeuGlyYrGlyAlaProProAlaGlyYr 157
Db 6845 GCTTCACCTGTGCGATATGAGCCCACTCTGATACGAGGCCCACTGCGAGATAT 6904
QY 158 GlYAlaProProLeuGlyYrGlyAlaProProLeuGlyYrGlyThrProProLeuGly 177
Db 6905 GAGGCCCACTCTGAGATATGAGCCCACTCTGATATGAGCCCACTCTCGGA 6964
QY 178 TyrGlyAlaProProLeuGlyYrGlyAlaProProAlaGlyAsnGlyYrProProAla 197
Db 6965 TATGAGGCCCACTCTGATATGAGCCCACTGCGAGAAATGAGGCCCTCG 7024
QY 198 GlYrYrAgaAlaSerProAlaGlySerGlyAlaArgProGlnGlySerThrAlaAlaGln 217
Db 7025 GGAATACAGAGCTCACTGCTGAGATCAGAGCCAGGCTCAGATCTACAGAGCCAG 7084
QY 218 AlaProGluAanglu 222
Db 7085 GCTCTGAAAACGAG 7099

RESULT 4
US-10-029-386-24981
; Sequence 24981, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: A60MICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24981
; LENGTH: 511

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; OTHER INFORMATION: MAP TO 297205.1
US-10-029-386-24981

Alignment Scores:
Pred. No.: 3.28e-16 Length: 511
Score: 256.00 Matches: 58
Percent Similarity: 47.14% Conservative: 8
Best Local Similarity: 41.43% Mismatches: 68
Query Match: 20.02% Indels: 6
DB: 13 Gaps: 2

US-09-864-291-12 (1-235) x US-10-029-386-24981 (1-511)
QY 102 GlYAlaProProAlaGlyYrGlyAlaProProGlyYrGlyYrAlaProPro 119
Db 90 GGTCTCCCTCTCTCACTGAGAGGCCCTCTCTCACTGAGAGTCCCTCTCTCACT 149
QY 120 AlaGlyYrGlyAlaGlnProValGlyAsnGlyYrProProValGlyYrArgAlaSer 139
Db 150 GCAAGGCTCCCTCTCACTGAGAGGCTCCCTCTCACTGAGAGGCTCCCTCTCA 209
QY 140 ProValArgYrGlyAlaProProLeuGlyYrGlyAlaProProAlaGlyYrGlyAla 159
Db 210 CCTGAGAGCTCCCTCTCACTGAGAGGCTCCCTCTCACTGAGAGGCTCCCTCT 269
QY 160 ProProLeuGlyYrGlyAlaProProLeuGlyYrGlyThrProProLeuGlyYrGly 179
Db 270 CCACCTGAGAGGCTCTCTCTCACTGAGAGGCTCTCTCTCACTGAGAGGCTCTCT 329
QY 180 AlaProProLeuGlyYrGlyAlaProProAlaGlyAsnGlyYrProProAlaGlyYr 199
Db 330 CTTCACTGAGAGGCTCTCTCTCACTGAGAGGCTCTCTCTCACTGAGAGGCTCT 389
QY 200 ArgAlaSerProAlaGlySerGlyAlaArgProGlnGlySerThrAlaAlaGlnAla 218
Db 390 TCTTCTCACTGAGAGGCTCTCTCTCACTGAGAGGCTCTCTCTCACTGAGAGG 449
QY 219 -----ProGluAangluAlaSerLeuProSerAlaSerSerGlnValAlaSer 235
Db 450 TCCCTCTCTCACTGAGAGGCTCTCTCTCACTGAGAGGCTCTCTCTCACTGAG 509

RESULT 5
US-09-864-761-15029
; Sequence 15029, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: A60MICA-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15029
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO Z99716.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
US-09-864-761-15029

Alignment Scores:
Pred. No.: 3,48e-15 Length: 471
Score: 245.00 Matches: 43
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19,16% Indels: 0
DB: 9 Gaps: 0

US-09-864-291-12 (1-235) x US-09-864-761-15029 (1-471)
QY 98 ValIleValTYrGlyAlaProProAlaGlyTYrGlyAlaProProGlyTYrGlyAla 117
DB 342 GTTATGTCCTATGAGAGCCCACTGCGAGATATGAGCCCACTCCCGATACGAGCC 401
QY 118 ProProAlaGlyTYrGlyAlaGlnProValGlyAsnGluGlyProProValGlyTYrG 137
DB 402 CCACTGCGAGATATGAGAGCCCACTGCGAGATATGAGAGCCCACTCCCGATACGAG 461
QY 138 AlaSerPro 140
DB 462 GCCTCACCT 470

RESULT 6
US-10-029-386-24981/c
; Sequence 24981, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
```

```

; SEQ ID NO 24981
; LENGTH: 511
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO Z97205.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
US-10-029-386-24981

Alignment Scores:
Pred. No.: 2,61e-13 Length: 511
Score: 226.00 Matches: 49
Percent Similarity: 45.87% Conservative: 1
Best Local Similarity: 44.95% Mismatches: 59
Query Match: 17,67% Indels: 0
DB: 13 Gaps: 0

US-09-864-291-12 (1-235) x US-10-029-386-24981 (1-511)
QY 102 GtYAlaProProAlaGlyTYrGlyAlaProProGlyTYrGlyAlaProProAlaGly 121
DB 499 GGAGAGAGAGCTGCGAGGTGAGAGGAGAGCTGCGAGGTGAGAGGAGAGCTGCGAGGT 440
QY 122 TyrGlyAlaGlnProValGlyAsnGluGlyProProValGlyTYrArgAlaSerProVal 141
DB 439 GGAGAGAGAGAGCTGCGAGGTGAGAGGAGAGCTGCGAGGTGAGAGAGAGAGCTGCA 380
QY 142 ArgTYrGlyAlaProProLeuGlyTYrGlyAlaProProAlaGlyTYrGlyAlaProPro 161
DB 379 GGTGAGAGAGAGAGCTGCGAGGTGAGAGGAGAGAGCTGCGAGGTGAGAGAGAGAGCT 320
QY 162 LeuGlyTYrGlyAlaProProLeuGlyTYrGlyThrProProLeuGlyTYrGlyAlaPro 181
DB 319 GCAAGTGAAGAGAGAGAGCTGCGAGGTGAGAGGAGAGAGCTGCGAGGTGAGAGAGAG 260
QY 182 ProLeuGlyTYrGlyAlaProProAlaGlyAsnGluGlyProProAlaGlyTYrArgAla 201
DB 259 CCTGAGGTGAGAGAGAGAGCTGCGAGGTGAGAGGAGAGAGCTGCGAGGTGAGAGAGG 200
QY 202 SerProAlaGlySerGlyAlaArgPro 210
DB 199 GAGCTGCGAGGTGAGAGAGAGAGCT 173

RESULT 7
US-10-029-386-11265
; Sequence 11265, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11265
; LENGTH: 541
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO Z97205.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
; OTHER INFORMATION: SWISSPROT HIT: P22793, EVALU8 1,00e-04
US-10-029-386-11265

Alignment Scores:
Pred. No.: 7,82e-12 Length: 541
Score: 211.00 Matches: 53
Percent Similarity: 45.86% Conservative: 8
Best Local Similarity: 39.85% Mismatches: 48
```

Query Match: 16.50% Indels: 24
DB: 13 Gaps: 5

US-09-864-291-12 (1-235) x US-10-029-386-11265 (1-541)

```

QY 109 G1YAlaProProGlyTyrGlyAlaProAlaGlyTyrGlyAlaGlyProValGly 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4 GGCCTCCCTCCCT-----CCACTCGAGGCTCTCTCTCCCTCCCTCGAGGC 48
QY 129 AaAGluGlyProProValGlyTyrArgAlaSerProValArgTyrGlyAlaProProleu 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 49 TCCCTCTCCCTCGAGGC-----TCTCTCT-----CAACTCGA 87
QY 149 G1YTyGlyAlaProProAlaGlyTyrGlyAlaProProleuGlyTyrGlyAlaProPro 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88 GGCCTCTCCCTCCCTCGAGGCTCTCTCTCCCTCGAGGCTCTCTCTCTCACT 147
QY 169 LeuGlyTyrGlyThrProProleuGlyTyrGlyAlaProProleuGlyTyrGlyAlaPro 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 148 GAGGCTCCCTCCCTCGAGGCTCTCTCTCCCTCGAGGCTCTCTCTCACT 207
QY 189 ProAlaGlyAaAGluGlyProProAlaGlyTyrArgAlaSerProAlaGlySerGlyAla 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 208 CCGAGGCTCCCTCCCTCCCTCGAGGCTCTCTCTCCCTCGAGGCTCTCTCTCACT 267
QY 209 ArgProGluGlySerThrAlaAlaGlyAlaProGluAaAGlu----- 222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 268 ---CCCGAGTCACTACCTGCTGCTCTCTCTCTCGAGGCTCTCTCGAGGCTCTCTCGA 324
QY 223 -----AlaSerIeuProSerIaSer 230
DB 325 CCAGGAGTCTCAATCTGTTCTCTCTCGAGGCTCTCT 363

```

RESULT 8

```

US-09-887-576-788
/ Sequence 788, Application US/09887576
/ Patent No. US20020144047A1
/ GENERAL INFORMATION:
/ APPLICANT: Budworth, P.
/ APPLICANT: Brown, D.
/ APPLICANT: Chang, H.
/ APPLICANT: Zhu, T.
/ APPLICANT: Han, B.
/ APPLICANT: Wang, X.
/ APPLICANT: Cooper, Bret
/ TITLE OF INVENTION: Promoters for regulation of plant expression
/ FILE REFERENCE: 1360.001US1
/ CURRENT APPLICATION NUMBER: US/09/887,576
/ CURRENT FILING DATE: 2001-06-25
/ PRIOR APPLICATION NUMBER: US 60/213,848
/ PRIOR FILING DATE: 2000-06-23
/ PRIOR APPLICATION NUMBER: US 60/214,087
/ PRIOR FILING DATE: 2000-06-23
/ PRIOR APPLICATION NUMBER: US 60/258,692
/ PRIOR FILING DATE: 2000-12-29
/ NUMBER OF SEQ ID NOS: 875
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 788
/ LENGTH: 1743
/ TYPE: DNA
/ ORGANISM: Oryza sativa
US-09-887-576-788

```

Alignment Scores:

```

Pred. No.: 1,79e-10 Length: 1743
Score: 202.50 Matches: 78
Percent Similarity: 38.89% Conservative: 13
Best Local Similarity: 33.33% Mismatches: 74
Query Match: 15.83% Indels: 70
DB: 10 Gaps: 13

```

US-09-864-291-12 (1-235) x US-09-887-576-788 (1-1743)

```

QY 54 LeuMetValIyBaAlaSerAlaValAlaArgGlyPheProIeuArgThrIeuAaASP 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1073 ATATATACAGAGGCTATAGAACTTCTGCTATAGGTCTCGCGCTAAT-TCACCTCAGAGAT 1131
QY 74 TrpPheSerIeuMetGlyIleTyrValIleThrGlyGluGlyAaMetCysThrProGln 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1132 ---GGTCCGACCAACACTATAGGTCTCAAGGTCTAGCTGTCAAGCATCTCTCGGG 1185
QY 94 MetProCysSerValIleValTyrGlyAlaProProAlaGlyTyrGlyAlaProProPro 113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1186 CAGCAAACTTCACTCTGCTGTGAGGACACT---GGTATTCAGCCCAACACT 1242
QY 114 -----GlyTyrGlyAla 117
DB 1243 ACTAGCCCTCTCAACTACCCGATCGAAGGCTCTCTCACTGCTGATATAGGCGC 1302
QY 118 ProPro-----AlaGlyTyrGlyAlaGlnProValGlyAaAGluGlyProProValGly 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1303 CCACCAACCAAGACTGCTATAGTACCACCC-----CAACCAAGAGAGG 1350
QY 136 TyrArgAlaSerProValArgTyrGlyAlaProProleuGlyTyrGlyAlaProPro--- 154
DB 1351 TAT-----AGTCAGGCGAGCTATAGGCACTTCGAGGAGCAAGAGGCTCTCTCAAC 1404
QY 155 ---AlaGlyTyrGly---AlaProProleu-----GlyTyrGlyAlaPro 167
DB 1405 ACTTCTCTTACGAGAGGAGGCGGCTCTCTGATCTCTCTGATGATAGG--- 1458
QY 168 ProleuGlyTyrGlyThrProProleuGlyTyrGlyAlaProProleuGlyTyrGlyAla 187
DB 1459 CAGTATGTTACTCTCAAGAACCAAGGCTATAGGCACTTCACCTTACCTCGTGCA 1518
QY 188 ProProAla----- 190
DB 1519 CCTGCTCAAGCCACCAAGGCTATAGTCAAGAGACTTATAGTATCTTATAGGAGT 1578
QY 191 GlyAaAGluGlyProProAlaGlyTyr----- 199
DB 1579 GAAAGCTACGGGAGGCTCGGCAATTTCTACTGAAGCTACACTGCTCTCTCCAG 1638
QY 200 -----ArgAlaSerProAlaGlySerGlyAlaArgProGluGlySerThr---Ala 215
DB 1639 GACCAATCTGCTTCGCGCCCTCGAGCTGCTGCGCTGCAACACGCTGCTCGGCA 1698
QY 216 AlaGlnAlaProGluAaAGluGlyAlaArgProGluGlySer 229
DB 1699 CCTACTGCTCTGAGAACAGTGAAGCCCAAGCCCTGCTACT 1740

```

RESULT 9

```

US-10-156-761-2531
/ Sequence 2531, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ CURRENT FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 2531
/ LENGTH: 2430
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis

```

```

FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2430)
US-10-156-761-2531

Alignment Scores:
Pred. No.: 7,73e-10 Length: 2430
Score: 197.50 Matches: 70
Percent Similarity: 38.89% Conservative: 7
Best Local Similarity: 35.35% Mismatches: 59
Query Match: 15,44% Indels: 62
DB: 15 Gaps: 11

US-09-864-291-12 (1-235) x US-10-156-761-2531 (1-2430)

QY 92 ProGlnMetProCyA-----SerVal11leValTYrGly----- 102
Db 367 CCAAGATCCCGTCCCGCCCGCGCGCCCAAGCCCGCGCGCGGTCAACCGCGCCCG 426
QY 103 ---AlaProProAlaGlyTYrGlyAlaProProGlyTYrGlyAlaProProAla--- 120
Db 427 GCGGACCGGAGGCGGCTACGGCTTCCGAGCCCGAGTCCGACTCCCGCGCGCC 486
QY 121 -----GlyTYr-----GlyAlaGlnProValGly 128
Db 487 GACTCCCGTCAAGCAGATGGTTACGGGTCCCGAGCGCGGTGCTCCAGGGCGTTCCG 546
QY 129 AsnGlnGlyProProValGlyTYrArgAlaSerProValArgTYrGlyAlaProProLeu 148
Db 547 CAAGCGGCGCCCGCGCGCGC---CCCATCTACCCGACGCCCAAGCGCGCCCGC----- 597
QY 149 GlyTYrGlyAlaProProAla-----GlyTYrGlyAla 159
Db 598 -----GCACTCTCCCGCGGCTCCAGCGCGCCCACTCCAGCGCGGTACGGCTTC 648
QY 160 ProProLeuGlyTYrGlyAlaProPro---LeuGlyTYrGlyThrProProLeuGlyTYr 178
Db 649 CCGGACCGCGCGGACCGCTTCCCGCCCAAGCGGCTTCCCGCAGCAGGGGCT 708
QY 179 -----GlyAlaProProLeuGlyTYrGly 186
Db 709 CCGGCTCCGGGCGCGCGCCCAAGACCAACCGCGCGCCCGCGGTACGGC 768
QY 187 -----AlaProProAlaGlyAsnGlnGlyProProAla----- 197
Db 769 TTCCCGAGCGCGCGCGCGCGCGAGTCTGTATCGGCAACCGCGCCACCGCGCAG 828
QY 198 ---GlyTYrArgAlaSerProAlaGlySerGlyAlaArgProGln-GluSerThrAla1 216
Db 829 CCGGCTACGGCTTCCCGCAGCAGAGGCCAGCGCGCGCCCGCAACCGCGGTACGGCTTC 888
QY 216 aglnAlaProGluAsnGlnAlaSerLeuProSerAlaSerSerGlnVal 233
Db 889 CCGGACCGCGCGCGCGCAGCAGCCAGCGCCGCGGCGCGCGAGGCTC 940

RESULT 10
US-10-156-761-1
Sequence 1, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156, 761
PRIOR FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30

```

```

PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: m1ac feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 4.61e-06 Length: 9025608
Score: 197.50 Matches: 70
Percent Similarity: 38.89% Conservative: 7
Best Local Similarity: 35.35% Mismatches: 59
Query Match: 15,44% Indels: 62
DB: 15 Gaps: 11

US-09-864-291-12 (1-235) x US-10-156-761-1 (1-9025608)

QY 92 ProGlnMetProCyA-----SerVal11leValTYrGly----- 102
Db 3119069 CCAAGATCCCGTCCCGCCCGCGCGCCCAAGCCCGCGCGGTCAACCGCGCCCG 3119128
QY 103 ---AlaProProAlaGlyTYrGlyAlaProProGlyTYrGlyAlaProProAla--- 120
Db 3119129 GCGGACCGGAGGCGGCTTCCCGCAGCGCGGAGTCCGACTCCCGCGCGCC 3119188
QY 121 -----GlyTYr-----GlyAlaGlnProValGly 128
Db 3119189 GACTCCCGTCAAGCAGATGGTTACGGGTCCCGAGCGCGGTGCTCCAGGGCGTTCCG 3119248
QY 129 AsnGlnGlyProProValGlyTYrArgAlaSerProValArgTYrGlyAlaProProLeu 148
Db 3119249 CAAGCGGCGCCCGCGCGCGC---CCCATCTACCCGACGCCCAAGCGCGCCCGC----- 3119299
QY 149 GlyTYrGlyAlaProProAla-----GlyTYrGlyAla 159
Db 3119300 -----GCACTCTCCCGCGGCTCCAGCGCGCCCACTCCAGCGCGGTACGGCTTC 3119350
QY 160 ProProLeuGlyTYrGlyAlaProPro---LeuGlyTYrGlyThrProProLeuGlyTYr 178
Db 3119351 CCGGACCGCGCGGACCGCTTCCCGCCCAAGCGGCTTCCCGCAGCAGGGGCT 3119410
QY 179 -----GlyAlaProProLeuGlyTYrGly 186
Db 3119411 CCGGCTCCGGGCGCGCGCCCAAGACCAACCGCGCGCCCGCGGTACGGC 3119470
QY 187 -----AlaProProAlaGlyAsnGlnGlyProProAla----- 197
Db 3119471 TTCCCGAGCGCGCGCGCGCGCGAGTCTGTATCGGCAACCGCGCCACCGCGCAG 3119530
QY 198 ---GlyTYrArgAlaSerProAlaGlySerGlyAlaArgProGln-GluSerThrAla1 216
Db 3119531 CCGGCTACGGCTTCCCGCAGCAGAGGCCAGCGCGCGCCCGCAACCGCGGTACGGCTTC 3119590
QY 216 aglnAlaProGluAsnGlnAlaSerLeuProSerAlaSerSerGlnVal 233
Db 3119591 CCGGACCGCGCGCGCGCAGCAGCCAGCGCCGCGGCGCGCGAGGCTC 3119642

RESULT 11
US-10-029-386-22839/c
Sequence 22839, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
EXPRESSION ANALYSIS TWO

```

```

FILE REFERENCE: ABOICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 22839
LENGTH: 866
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005630.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
OTHER INFORMATION: SWISSPROT HIT: Q08379, EVALUE 4.00e-10
OTHER INFORMATION: NT HIT: G1161055, EVALUE 4.00e-95
OTHER INFORMATION: EST_HUMAN HIT: BF828725.1, EVALUE 0.00e+00
US-10-029-386-22839

Alignment Scores:
Pred. No.: 7,91e-10 Length: 866
Score: 192.50 Matches: 45
Percent Similarity: 47.59% Conservative: 24
Best Local Similarity: 31.03% Mismatches: 73
Query Match: 15.05% Indels: 3
DB: 13 Gaps: 3

US-09-864-291-12 (1-235) x US-10-029-386-22839 (1-866)
QY 91 ThrProGlnMetProGlySerValIleValIYrGlyAlaProProAlaGlyYrGlyAla 110
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
600 TCTCCGCGCCCTGACAGCTCTCTCTCTGTCACACAGCTCTCTCTCTGTCACATGAGCTCT 541
QY 111 ProProProGlyYrGlyAlaProProAlaGlyYrGlyAlaGlnProValGlyAlaGln 130
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
540 CTTCCGTCACACAGCTCTCTCTCTGTCACATGAGCTCTCTCTCTGTCACACAGCC 481
QY 131 GlyProProValGlyYrGlyAlaSerProValGlyYrGlyAlaProProGlnGlyYr 150
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 TCTCCCTCTGTCACAGCTCTCTCTCTGTCACACAGCTCTCTCTCTGTCACATG 421
QY 151 GlyAlaProProAlaGlyYrGlyAlaProProGlnGlyYrGlyAlaProProGln 170
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
420 GCTCTCTCTCTGTCACAGCTCTCTCTCTGTCACATGAGCTCTCTCTCTGTCAC 361
QY 171 YrGlyThrProProGlnGlyYrGlyAlaProProGlnGlyYrGlyAlaProProAla 190
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 GATGCTCTCTCTCTGTCACATGAGCTCTCTCTCTGTCACATGAGCTCTCTCT 301
QY 191 GlyAlaGlnGlyYrProProAlaGlyYrGlyAlaSerProAlaGlySerGlyAlaG 210
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300 CAGTACGCTCTCTCTCTGTCACATGAGCTCTCTCTCTGTCACATGAGCTCTCT 244
QY 211 GlnGlnSerThrAlaAlaGlnAlaProGlnGlnGlnAlaSerLeuPro--SerAlaSer 229
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
243 CCTGTCACACAGCTCT--CCTCTGTCATGATGAGCTCTCTCTCTGTCATGATG 187
QY 230 SerSerGlnValHis 234
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 172

RESULT 12
US-09-764-891-1356
Sequence 1356, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231

```

```

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1356
LENGTH: 321
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (316)
OTHER INFORMATION: n equals a,t,g, or c
LOCATION: (321)
OTHER INFORMATION: (321)
US-09-764-891-1356

Alignment Scores:
Pred. No.: 3.09e-10 Length: 321
Score: 192.00 Matches: 39
Percent Similarity: 95.12% Conservative: 0
Best Local Similarity: 95.12% Mismatches: 1
Query Match: 15.01% Indels: 1
DB: 11 Gaps: 0

US-09-864-291-12 (1-235) x US-09-764-891-1356 (1-321)
QY 1 MetProPheApleMetThrAsnLeuThrValGlnGlnProValPheAlaAsn 20
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
195 ATGCCATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 254
QY 21 IleYsGlyThrIleGlnAlaAlaProYrGlyGlyYrGlyAlaAlaThrPhe 40
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 ATTAAGGACATTAATGACGAGCTCTCATATGAGCTCTGGAAGGACAAAGCTTATA 314
QY 40 P 40
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
315 A 315

RESULT 13
US-09-918-995-33536
Sequence 33536, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 33536
LENGTH: 432
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)-(432)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-33536

Alignment Scores:
Pred. No.: 5.29e-10 Length: 432
Score: 191.00 Matches: 36
Percent Similarity: 72.31% Conservative: 11
Best Local Similarity: 55.38% Mismatches: 18
Query Match: 14.93% Indels: 0
DB: 11 Gaps: 0

US-09-864-291-12 (1-235) x US-09-918-995-33536 (1-432)
QY 1 MetProPheApleMetThrAsnLeuThrValGlnGlnProValPheAlaAsn 20
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db 231 ATGCATTATCTATCATGAAAGACTGTGATCAAGACCCCGATTGTGCAAACTAC 230
 Qy 21 ILELYGLYTHRIIEGLNALAIAProTYRGLYGLYTHRIIUGLYGLNALIAIThrPheLYe 40
 Db 291 ATCAAGGGAACAGAGAGGGCGAGAGCGGAGGTGGCTGGAGAGCTCTGCTCTCAAG 350
 Qy 41 LeuValPheArgAnGIYAaPAlaIleGLIuPheAlaInleuMetValLYsaIaIaSer 60
 Db 351 TTGACTTCAAGGAGGGGGCGCATTTGATTGGAGAGGAGATGCTCCAGGTGGCATCT 410
 Qy 61 AlavaIaIaArgLY 65
 Db 411 CAAGCTCCAGAGGT 425
 RESULT 14
 US-10-029-386-20220
 ; Sequence 20220, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
 ; FILE REFERENCE: AECOMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 20220
 ; LENGTH: 780
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC005263.1
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.5
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
 ; OTHER INFORMATION: EST HUMAN HIT: AL539431.1, EVALUATE 0.00e+00
 ; OTHER INFORMATION: SWISSPROT HIT: P04280, EVALUATE 1.00e-04
 ; OTHER INFORMATION: NT HIT: g115307287, EVALUATE 0.00e+00
 ; US-10-029-386-20220
 Alignment Scores:
 Pred. No.: 3.76e-09 Length: 780
 Score: 185.00 Matches: 58
 Percent Similarity: 38.86% Conservative: 17
 Best Local Similarity: 30.05% Mismatches: 67
 Query Match: 14.46% Indels: 51
 Gaps: 7
 US-09-864-291-12 (1-235) x US-10-029-386-20220 (1-780)
 Qy 92 ProGIInetProCySerValIleValTYRGLY-----AlaProProAlaGLYTYR 108
 Db 180 CCCACAGAGGCTCGCCCTCAGAGGCCCGGAGACACCCAGCTACCCCGCAGCTCC 239
 Qy 109 GLYAlaProProProGLY---TYRGLYAlaProProAlaGLYTYRGLYAlaInProVal 127
 Db 240 AGGGGTCCACCCCGGCGCCAGGTGGTCATCCCTGTCATCTGGGGTCCATCCCCAGC 299
 Qy 128 GLYAnGIUGLYProProValIGLYTYRArgAlaSerProValArgTYRGLYAlaProPro 147
 Db 300 TCTGGGCTCCACCCCGCAGCTCTGGGTCTCATCTCCACGCCCTGGGTCCACCCAGC 359
 Qy 148 LeuGLYTYRGLYAlaProProAlaGLYTYRGLYAlaPro-----ProLeuGLY----- 163
 Db 360 AACCTGGGGTCCACCCCGCAGCTCTGAGTCCACCTCCAGGCCCGGGGGTTCAACC 419

Qy 164 -----TYRGLYAlaPro----- 167
 Db 420 ACCAGCCCCGAGTCCACCCACAGCCCTGGGGTTCAACCCAGCCAGGGGTCA 479
 Qy 168 -----ProLeuGLYTYRGLYThrProProLeuGLYTYRGLYAlaPro----- 181
 Db 480 TCTCTCCCATCAGCGGGGGTTCAACCCAGGCCCGGGGGTGCACCCAGAGCCCCC 539
 Qy 182 -----ProLeuGLYTYR 185
 Db 540 CATTACCTCCAGAGGCCCGGAGGTGCACCCAGCCCGAGGATTCACCTCAGGCC 599
 Qy 186 GLYAlaProProAlaGLYAnGIUGLYProProAlaGLYTYRArgAlaSerProAlaGLY 205
 Db 600 GGGGGTCCACCCCGCAGCTCCCGGGGTCCATCGTGGCTCTGGGGTCCACCTCAGCC 659
 Qy 206 SerGLYAlaArgProGIInleuSerThrAlaIa-----GlnAlaProGIInleu 221
 Db 660 TCCGGAGTTCACCCCTCAATCTGGGGTGCACCCCGCACTCCCATGCTCCATGCT 719
 Qy 222 GLUAla-SerLeuProSerAlaSerSerGIInVal 233
 Db 720 GAGGCCCCCATCTCCCTCCGAAAGGCCAGGAAACATA 756
 RESULT 15
 US-09-938-842A-2004
 ; Sequence 2004, Application US/09938842A
 ; Patent No. US20020160378A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Kreps, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; TITLE OF INVENTION: SAME, AND METHODS OF USE
 ; FILE REFERENCE: SRIPI300-3
 ; CURRENT APPLICATION NUMBER: US/09/938,842A
 ; CURRENT FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 60/227,866
 ; PRIOR FILING DATE: 2000-08-24
 ; PRIOR APPLICATION NUMBER: US 60/264,647
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/300,111
 ; PRIOR FILING DATE: 2001-06-22
 ; NUMBER OF SEQ ID NOS: 5379
 ; SEQ ID NO 2004
 ; LENGTH: 2016
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-938-842A-2004
 Alignment Scores:
 Pred. No.: 1.03e-08 Length: 2016
 Score: 185.00 Matches: 47
 Percent Similarity: 41.35% Conservative: 8
 Best Local Similarity: 35.34% Mismatches: 40
 Query Match: 14.46% Indels: 38
 Gaps: 6
 US-09-864-291-12 (1-235) x US-09-938-842A-2004 (1-2016)
 Qy 101 TYRGLYAlaProProAlaGLYTYRGLYAlaProProProGLYTYRGLYAla----- 117
 Db 355 TTGGGGGTCCAGAGGAATAT-----CCACATATTAATTGGGCTATCGATGAG 405
 Qy 118 -----ProProAlaGLYTYRGLYAlaInProValIGLYAsn 129
 Db 406 TTAAGGCTTGTGAAGTGAATCCCGCGTTCACGTGATGCTCCCTATTATTAACA 465
 Qy 130 GlnGLYProProValIGLYTYRArgAlaSerProValArg-----TYRGLY 144
 Db 466 -----CCACAGTGAAGTCA-----CCACGTTAAGACGCTCAGGGGGCTATGGA 513

Qy 145 AlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaProProLeuGlyTyr 164
Db 514 GTACCTCCAGCTGGCTATGAGATCAAGCTGGTTATGAGATCAAGCGGTTAT 573
Qy 165 GlyAlaProProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAlaProProLeuGly 184
Db 574 GGAATACCTCAAGCTGGCTATGAGATCAAGCGGTTATGAGATCAAGCTGGCT 633
Qy 185 TyrGlyAlaProProAlaGlyAangGlyProProAlaGlyTyrArgAlaSerProAla 204
Db 634 TAGGA-----ATGCCAGT 648
Qy 205 GlySerGlyAlaArgProGlnGlnSerThrAlaAlaGln 217
Db 649 GGTTCCTCGAAGAGCTTGATGAAGCATGCGAATGAG 687

Search completed: December 17, 2003, 01:19:21
Job time : 15959.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 17:31:55 (Search time 1518.07 Seconds)

(without alignments)
3762.388 Million cell updates/sec

Title: US-09-864-291-12
Perfect score: 1379
Sequence: 1 MPPDLMNLTVEQVFPANP.....AQAPENASLPSSSSQVHS 235

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Pgapop 6.0, Pgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2.1/USPRO.spool/p/US09864291/runat.15122003.160859.23318/app.query.fasta.1.846
-DB=EST -OPWT=fastest -SUFFIX=est -MINMATCH=0.1 -LOOCL=0 -LOOEXT=0
-UNITS=bite -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human0.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZB=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09864291.QCGN.1.1.3724 @runat.15122003.160859.23318 -NCPU=6 -ICPU=3
-NO MAP -LARGESUBSTRY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEFTIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Pgapop=6
-Pgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1207	94.4	1079	12	BMS64167
2	755.5	53.0	1414	11	AK015863
3	677.5	59.0	775	10	BG722815
4	625	48.9	663	10	BG699398
5	511.5	40.0	717	14	BT15414
6	468	36.6	701	10	BG701881
7	350	27.4	488	28	AQ997923
8	347.5	27.2	814	13	BU961805
9	340.5	26.6	931	13	BU914288
10	338.5	26.5	744	13	BU230211
11	338	26.4	932	13	BU107273
12	323.5	25.3	892	14	CD301781
13	319	24.9	816	13	BU916890
14	318	24.9	868	14	CD255258
15	312	24.4	862	13	BQ735296
16	311.5	24.4	426	28	AQ997923
17	306	23.9	636	12	BM490729
18	306	23.9	643	9	AL868685
19	305	23.8	723	9	AJ452591
20	299.5	23.4	719	14	CB593075
21	294	23.0	700	13	BU284977
22	290.5	22.7	655	9	AL898723
23	290.5	22.7	661	9	AL972453
24	290.5	22.7	663	9	AL896682
25	290	22.7	647	9	AL849651
26	287	22.4	618	13	BU806987
27	287	22.4	738	13	BX082055
28	285	22.3	663	9	AL881810
29	282.5	22.1	594	12	BM190972
30	282.5	22.1	735	14	CA364472
31	282	22.0	607	12	BM190975
32	281	22.0	635	9	AL889954
33	280.5	21.9	568	12	BI390687
34	279	21.8	1004	13	BQ217128
35	279	21.8	1120	14	CD499460
36	277.5	21.7	829	13	BU413257
37	273.5	21.4	584	12	BJ501828
38	272	21.3	1028	13	BU541892
39	271	21.2	747	13	BQ229609
40	270.5	21.1	922	13	BQ645308
41	269.5	21.1	718	14	CB591312
42	269.5	21.1	776	13	BU846529
43	269.5	21.1	878	13	BQ723331
44	269.5	21.1	913	13	BQ718987
45	269.5	21.1	937	13	BU174627

ALIGNMENTS

RESULT 1
BMS64167
LOCUS
DEFINITION AGNCOURT 6560149 NIH_MGC_119 Homo sapiens CDNA clone IMAGE:5742129
5', mRNA sequence.
ACCESSION
BMS64167
VERSION
BMS64167.1 GI:18811738
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1079)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LHAM12759 Row: m Column: 10
 High quality sequence stop: 726.
 Location/Qualifiers

FEATURES
 source
 1. 1079
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5742129"
 /rname="cDNA:5742129"
 /rsize="1108"
 /lab_host="DH10B"
 /clone_11b="NIH MGC 119"
 /note="Organ: brain; Vector: pCMV-Sport6; Site: 1: NotI;
 Site 2: EcoRV (destroyed); RNA source: normal medulla from
 anonymous male age 27. Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.3 kb, insert size range
 0.9-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 013. Note:
 this is a NIH MGC library."

BASE COUNT 267 a 297 c 268 g 247 t
ORIGIN

Alignment Scores:
 Pred. No.: 1.47e-70 Length: 1079
 Score: 1207.00 Matches: 229
 Percent Similarity: 97.06% Conservative: 2
 Best Local Similarity: 96.22% Mismatches: 4
 Query Match: 94.37% Indels: 3
 DB: 12 Gaps: 0

US-09-864-291-12 (1-235) x BM564167 (1-1079)

Qy 1 MeProheapleuMetThrasLeuThraValGluInProValPheAlaAAspPhe 20
Db 269 ATGCCATTGATCTGATGACGACCTCACTTTGAACACGATTTGCTGCAACTTC 328
Qy 21 TlelyeglyThrlleglnalaaIProTyrglygIyTTProgluglgnalathrPheLys 40
Db 329 ATTAGGGAATATTCAAGGAGCTCCATATGCTGGCGGAGGACCAAGCTACTTTAAA 388
Qy 41 LeuValPheArGaenglyAspAlaIleglnPheAlaGlnLeuMetVallyPheAlaSer 60
Db 389 TTACTCTTCAGAAATGAGGTCGACATGAAATTCGCCAGTGAAGGGAAGGCTGCTCT 448
Qy 61 AlAvalAlaArggIyPheProLeuArGthrLeuAsnAspTTPheSerSerMetGlyIle 80
Db 449 GCTGCTGCCGAGGATTTCCACTTAAGACCTTAATGACTGTTCAAGCTCTATGGAAAT 508
Qy 81 TyValIleThrglygluglyAsnMetCyArThrProGlnMetProCySerValIleVal 100
Db 509 TATGTAATTACTGGGGAAGGAAATATGTCATCTCAAGATGCCCTGTTCAGTTATGTC 568
Qy 101 TyrglyAlaProProAlaGlyTyrglyAlaProProProGlyTyrglyAlaProProAla 120
Db 569 TATGAGGCCCACTCGAGATATGAGCCCACTCCGATATGAGGAGCCCACTCGCA 628
Qy 121 GlyTyrglyAlaGlnProValGlyAsnGlnGlyProProValGlyTyrglyAlaSerPro 140
Db 629 GGAATATGAGCCCACTCGATATGAGGAGCCCACTCGATATGAGGAGCCCACTCGCT 688

Qy 141 VALATGTYTGYVALAProProLeuGlyTYTGYVALAProProAlaGlyTYTGYVALA-Pr 160
Db 689 GTGCAATATGAGGCCCACTCTTGATGAGAGGCCCACTCGAGATATGAGGCCCC 748
Qy 160 oProLeuGlyTYTGYVALAProProLeuGlyTYTGYThrProProLeuGlyTYTGYVAL 180
Db 749 ACCTCTAGATATGAGGCCCACTCTTGATGAGAGGCCCACTCGATATGAGGCC 808
Qy 180 aProProLeuGlyTYTGYVALAProProAlaGlyAsnGlnGlyProProAlaGlyTYTGY 200
Db 809 CCCACTCTCGATATGAGGCCCACTCGAGAGGAGGCCCACTCGAGATATGAGGCC 868
Qy 200 gAlaSerProAlaGlySerGlyVALAArgProGlnGlnSerThrAla-AlaGlnAlaProG 220
Db 869 AGCTCACTGCTGATGAGAGGCCCACTCGAGAGGAGGCCCACTCGAGATATGAGGCC 928
Qy 220 lAAsnGlu-AlaSerLeuProSerAlaSerSerSerGlnValAlaSer 235
Db 929 GAAAGAGAGCTTCTTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 976

RESULT 2
AK015863
LOCUS
DEFINITION
 Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930521123 product:hypothetical protein, full insert sequence.

ACCESSION
AK015863
VERSION
AK015863.1 GI:12854363
KEYWORDS
HTC, CAP trapper.
SOURCE
 Mus musculus (house mouse)
ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
 Carminci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subcloning of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carminci, P., Konno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Harada, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Iizawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, D., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multiplexed capillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
 Kawal, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamahata, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Pfelechemann, W., Gaasterland, T., Giesel, C., Kling, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Tomita, M., Quackenbush, J., Schmitt, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Bareh, G., Blake, J., Boffelli, D., Bojunga, N., Carminci, P., de Bonaldo, M. P., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Guelinich, S., Hill, D.,

Db 991 GCCACCTCTGTACACAGAGACCTCCACCTGCTGCTGACAGACCATCCCATGCTC 1050
 Qy 214 ThrAlaAlaGlnAlaProGluAemGluAlaSerLeuProSerAlaSerSerGlnVal 233
 Db 1051 GTGGAGTCACCAATCTGAATTCACAGACATCTTTCTCTACCTGCTCCTCACAAGTC 1110
 Qy 234 HIsSer 235
 Db 1111 CACTCT 1116
 RESULT 3
 BG722815 775 bp mRNA linear EST 08-MAY-2001
 LOCUS 602695278F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4827427 5',
 DEFINITION mRNA sequence.
 ACCESSION BG722815
 VERSION BG722815.1 GI:14002002
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE NIH-MGC http://mgc.nci.nih.gov/
 1 (bases 1 to 775)
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroaki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM10743 row: 1 column: 20
 High quality sequence stop: 651.
 Location/Qualifiers
 1..775
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4827427"
 /lab_host="DH108"
 /clone_1lb="NIH_MGC_97"
 /note="Organ: testis; Vector: pBluescriptR (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVA-3',
 size selected for average insert size 2.2 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 191 a 212 c 193 g 179 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.01e-35 Length: 775
 Score: 677.50 Matches: 144
 Percent Similarity: 88.69% Conservativeness: 5
 Best Local Similarity: 85.71% Mismatch: 13
 Query Match: 52.97% Indels: 7
 DB: 10 Gaps: 1
 US-09-864-291-12 (1-235) x BG722815 (1-775)
 Qy 1 MetProPheAspLeuMetTrpAsnLeuThrValGluGlnProValIleAlaAlaAspPhe 20
 Db 276 ATGCCATTGATTCATGACGACACCTCAGCTTGAACACACAGATTGCTGCAAACTTC 335
 Qy 21 IleValGlyThrIleGlnAlaAlaProTyrGlyGlyTrpGluGlyGlnAlaThrPheLys 40

Db 336 ATTTAGGAACTATTTCAGCAAGCTCCATTCATTCGCTGCGAAAGACAACTCTTTTAA 395
 Qy 41 LeuValIlePheArgAsnGlyValAlaIleGluPheAlaGlnLeuMetValIleValAsp 60
 Db 396 TTAGCTTCAGAAATGAGATGCTGCATTTAAATTTGCGAGTTATGATGTAAGCTGCTCT 455
 Qy 61 AlAValAlAArgGlyPheProLeuAlaGlyThrLeuAsn-AspTrpPheSerSerMetGly 80
 Db 456 GCTGTGCTCCGAGAGATTCCTACCTTAGAACCTTAACTGACTGATTCATATGGAAT 515
 Qy 80 eTyrValIleThrGlyGlyGlyValAspMetCysThrProGlnMetProCysSerValIleVal 100
 Db 516 TTATGTAACTTACTGGGAAAGGAAATATGTGACTCCACAGATGCTTGTTCATATCT 575
 Qy 100 TTTGTAlAProProAlaGly-TTGTAlAProProProGlyTTGTAlAProProAla 120
 Db 576 CTATGAGGCCCAACGACGATATGAGAGCCCACTCCCGATATGAGAGC-CCAACTG 634
 Qy 120 TAglyTyrGlyAlaGlnProValGlyAlaGlnGlyProProValGlyTyrAlaAlaSer- 139
 Db 635 CAGGATATGAGACCCCAACCCGTAGAAACAGACGGCCGCTGTGGATACAGAGCTCTAA 694
 Qy 140 ProValArgTyr-GlyAlaProProLeuGly-TTGTAlAProProAlaGlyTyr---G 158
 Db 695 CTTGTCCAGCATGTAAGCCCACTCTCGGCAATACCGAGCCCACTGACGACATGG 754
 Qy 158 TAlAProProLeu 162
 Db 755 AAGCCCACTCTGA 768
 RESULT 4
 BG699398 663 bp mRNA linear EST 07-MAY-2001
 LOCUS 602679081F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4811804 5',
 DEFINITION mRNA sequence.
 ACCESSION BG699398
 VERSION BG699398.1 GI:13967653
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE NIH-MGC http://mgc.nci.nih.gov/
 1 (bases 1 to 663)
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroaki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM10703 row: a column: 21
 High quality sequence stop: 660.
 Location/Qualifiers
 1..663
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4811804"
 /issue_type="hipocampus"
 /lab_host="DH108"
 /clone_1lb="NIH_MGC_95"
 /note="Organ: Brain; Vector: pBluescriptR (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVA-3',
 size selected for average insert size 2.5 kb and
 normalized to ROT 5. This is a primary library enriched

for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 168 a 158 c 169 g 164 t 4 others

ALIGNMENT SCORES:

Score:	2.55e-32	Length:	663
Percent Similarity:	93.33%	Matches:	126
Best Local Similarity:	93.33%	Conservative:	0
Query Match:	48.87%	Mismatches:	9
		Indels:	2
		Gaps:	0

US-09-864-291-12 (1-235) x BG699398 (1-663)

QY 1 MetProPheAplemMetThraMetThraValGluGlnProValPheAlaAlaAse 20
 DB 258 ATGCCATTGATCTGATGACGAACTCTGATGACGAACTCTGATGACGAACTCTC 317
 QY 21 IleyGlyThrIleGlnAlaAlaProTyGlyTyTPGluGlyGlnAlaThrPhe 40
 DB 318 ATTAAGGAACTATTCAGGAGCTCATTCGTGCTGGGAAAGCAAGCTTCTTAA 377
 QY 41 LeuValPheArgAseGlyAsePalaileGluPheAlaGlnLeuMetVallyAlaAse 60
 DB 378 TTAGCTCTCAGAAATGAGAGGTCCCATGGAATTTGCCCATGATGATGAAAGCTCTCT 437
 QY 61 AlAValAlaAArgGlyPheProLeuArgThrLeuAseAAPTTPPheSerSerMetGly 80
 DB 438 GCTGCTGCCCGAGGATTCCTCATGACCTTAAATGACGTGTTAGCTTATGGGAAT 497
 QY 81 TyrValIleThrGluGluGlyAseMetCysThrProGlnMetProCysSerValIleVal 100
 DB 498 TATGTAATTAATCTGGGAAAGGAAATGTCACATCCACAGATCCTTGTTCAGTTATTC 557
 QY 101 TyTGlyAlaProProAlaGlyTyTyGlyAlaProProProGlyTyTyGlyAlaProProAla 120
 DB 558 TATGAGAGCCCACTGACAGATATGAGAGCCCACT-CCCGATACGAGAGCCCACT-GCA 615
 QY 121 GLYTYTYGlyAlaGlnProValGlyAseGlyTyProProValGly 135
 DB 616 GGAATYTGAGCCCANACCTGAGAAATGAAAGCGCGCTGTGGGA 660

RESULT 5 BY715414 717 bp mRNA linear EST 17-DEC-2002
 LOCUS BY715414 RIKEN full-length enriched, adult male testis Mus musculus
 DEFINITION BY715414 RIKEN full-length enriched, adult male testis Mus musculus
 CDNA clone 4930521123 5', mRNA sequence.
 ACCESSION BY715414
 VERSION BY715414.1 GI:27128531
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 717)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nishida, I., Oshio, N., Saito, R., Suzuki, H., Yamashita, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gotohori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schreml, L.M., Kanapin, A., Matsumura, H., Batalov, S., Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Choich, C., Corbani, L.B., Cousins, S., Datta, B., Dragan, T.A., Fletcher, C.P., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grigoriadis, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltale, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pereira, G., Pesole, G., Petrovsky, N., Pillai, R., Ponting, J.C., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reid, D.J., Ring

JOURNAL MEDLINE PUBLISHED

COMMENT

B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shinde, K., Sultana, R., Takemura, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyszynski, A., Yang, Y., Yang, Y., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Akawa, T., Fukuda, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 12466851
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Adachi, J., Aizawa, K., Akimura, T., Akawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Munazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tsgami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in RIKEN.
 Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES SOURCE

Location/Qualifiers
 1..717
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="4930521123"
 /sex="male"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="SOLR"
 /note="Site 1: XhoI; site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGGAGAGAGAGATCCAGAGCTTTTCTTTTCTTTTNN 3']. cDNA was prepared by using trihaloase thermo-activated reverse transcriptase and subsequently enriched for full-length by


```
Db          677 AGAGCTCACC-----GGCGCCCA 697
```

RESULT 7
LOCUS AQ97933/c
DEFINITION RPCT-23-384A20.TV RPCT-23 Mus musculus genomic clone RPCT-23-384A20
ACCESSION AQ97933
VERSION AQ97933.1 GI:7072953
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Zukarya; Muezaia; Chorata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 Zhao,S., Nieman,W., Feldbylum,T., Malek,J., Shatman,S., Akintet
 1 (bases 1 to 488)
 B., Levins,M., McGann,S., Tseaye,G., Geer,K., Krol,M., de Jong,P.
 and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCT-23
COMMENT Unpublished
 Other_GSAs: RPCT-23-384A20.TV
 Contact: Shayang Zhao
 Department of Bukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCT-23. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Resea ch Genetics (into@resgen.com). BAC end page:
 http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 384 row: A column: 20
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..488
 /organism="Mus musculus"
 /mol_type="Genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCT-23-384A20"
 /sex="Female"
 /lab_host="DH10B"
 /clone_1fb="RPCT-23"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1;
 Scori; Site: 2: Scori; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI MethyIase. Size
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)." " "

BASE COUNT 102 a 128 c 157 g 100 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 2,77e-14 Length: 488
Score: 350.00 Matches: 73
Best Similarity: 52.26% Conservative: 8
Percent Similarity: 47.10% Mismatches: 34
Query Match: 27.37% Indels: 40
DB: 28 Gaps: 4

US-09-864-291-12 (1-235) x AQ97933 (1-488)

CY 101 TGTGTAATAProProAlaGlyTYTGTAATAProProGlyTYTGTAATAProProAla 120

DB 466 TATTGGGGCTCACCGGTGGGAATACGAATGCCATTCGCAGGTGTATGAAAGCCCATCTCGG 407

CY 121 GTTTGTGTAAGAInProValGlYanGLuglYProPProValGlyTTYTaAglaaSerPro 140

Db		406	GAGTACGAGAGNCAGCCAGTGAGGTATGTAAGGCCCACTCCCGATTACGAATGTCCTGGCT	347
Oy		141	-----ValArgTYrGYAlaPro-----	Proleu 148
Db		346	CCTGGATACGAGAGCTGTAGATATGAGTGGCTCCTCCTCATATATGATACCCTAATG	287
Oy		149	GIYTYrGYAlaPProProAlaGIYTYrGYAlaPProProLeuGIYTYrGYAlaPProPro	168
Db		286	GAGCTATGAGACTTCACCTCCTGGATATAGAACCTTCACCTGTAGATACGAAATCCCACCT	227
Oy		169	LeuGIYTYr	171
Db		226	CTGGATATGTAGAAGCCCCCACTATGATATAGAGCCCACTCCCAATAAAAAACTACA	167
Oy		172	-----GIYThPProProLeuGIYTYrGYAlaPProProLeuGIYTYrGYAla	187
Db		166	CTATGAGATCCGAGTTCCCACTCCCACTATATTAACCCCGGCTATGAGGTATGAACCT	107
Oy		188	ProProAlaGIYAAnGIUdIYProProAlaGIYTYrArpaAlaser-----	202
Db		106	CCTCTCTTAGGCGCTAGAGTCCATACCTCTCGAATCAAAGGCACAGCTCTGACACAGAAA	47
Oy		203	-----ProAlaGIserGIYAAlaArgProGIInIuSerThrAla	215
Db		46	GCTCCACCTGCTGCTGCTCTGAAGACAGAGCACATCCCATGCTGTGGCA	2
RESULT 8				
BUS61805		814 bp	mRNA	linear EST 21-OCT-2002
LOCUS		AGNCOOURT 10617160 NIH_MGC_169	Mus musculus	cDNA clone
DEFINITION		IMAGR:6742447 5', mRNA sequence.		
ACCESSION		BUS61805		
VERSION		BUS61805.1	GI:24191377	
KEYWORDS		EST.		
SOURCE		Mus musculus	(house mouse)	
ORGANISM		Mus musculus		
REFERENCE		Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE		1 (bases 1 to 814)		
JOURNAL		NIH-MGC http://mgs.nci.nih.gov/.		
COMMENT		National Institutes of Health, Mammalian Gene Collection (MGC)		
		Unpublished		
		Contact: Robert Straubeberg, Ph.D.		
		Email: cgapds-remail.nih.gov		
		Tissue Procurement: Dr. Jonathan Kuo, NIMH		
		cDNA Library Preparation: Michael Brownstein Laboratory		
		cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
		DNA Sequencing by: Agencourt Bioscience Corporation		
		Clone distribution: MGC clone distribution information can be		
		found through the I.M.A.G.E. Consortium/LNL at:		
		http://image.llnl.gov		
		Plate: LLCM3080 row: 1 column: 06		
		High quality sequence stop: 504.		
FEATURES		Location/Qualifiers		
Source		1..814		
		/organism="Mus musculus"		
		/mol_type="mRNA"		
		/db_xref="taxon:10090"		
		/clone="IMAGR:6742447"		
		/lab_host="DH10B (TI-phase-resistant)"		
		/clone_id="NIH_MGC_169"		
		/note="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfiI		
		(ggcgcatgagcc)/ Site 2: SfiI (ggcgcatgagcc); cDNA made		
		by oligo-dT priming and directionally cloned. 5' and 3'		
		adaptors were used in cloning as follows:		
		5'-AMGCGTGTATCAACGAGAGTGGCAATGACATTCAGCCGAG-3' and		
		5'-ATTCTAGAGCCGAGGCGGCGGAGATG-dT(30)NN-3'. Full-length		
		enriched library was constructed using the Clontech		
		Creator SMART kit and size-selected to contain the 0.5 kb		
		size fraction. Library created in the laboratory of M.		
		Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."		
BASE COUNT		207 a	238 c	177 g
ORIGIN			190 t	2 others

BU230211
LOCUS BU230211 744 bp mRNA linear EST 26-NOV-2002
DEFINITION 603398308F1 CSBQCHN23 Gallus gallus cDNA clone CHEST28901 5', mRNA
SEQUENCE
ACCESSION BU230211
VERSION BU230211.1 GI:25470774
KEYWORDS EST
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Phasianidae; Gallus.
REFERENCE Boardman, P.B., Sanz-Bzquez, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.,
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
2235534
JOURNAL PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 10D, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES
Source location/Qualifiers
1..744
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST28901"
/dev_stage="22"
/lab_host="DH10B"
/clone_1ib="CSBQCHN23"
/note="Organ: heads; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."
BASE COUNT 164 a 225 c 179 g 176 t
ORIGIN
Alignment Scores:
Pred. No.: 2,386-13 Length: 744
Score: 338.50 Matches: 81
Percent Similarity: 44.93% Conservative: 12
Best Local Similarity: 39.13% Mismatches: 57
Query Match: 26.47% Indels: 57
DB: 13 Gaps: 6
US-09-864-291-12 (1-235) x BU230211 (1-744)
QY 1 MetProPheAspLeuMetThrAsnLeuThrValGluGlnProValPheAlaAlaAspHe 20
DB 45 ATGCCGTTTATTATTGTGAAGAGGCTCTATTGAGCGACCTTTTCCTGCTAATTAC 104
QY 21 TlelyGgIyThnIleGlnAlaAlaProTyrgIyGlyTTPGluGlyGlnAlaThrPheIys 40
DB 105 ATCAAGAAGACAGATTCAAGCTGAGCAAGAGGCTGCGAGGAGGAGGAGGAGGAGG 164
QY 41 LeuValPheArgAsnGlyAspAlaIleGluPheAlaGlnLeuMetValIysAlaIaIaser 60

DB 165 CTGACTTTCACACACGAGGAGGACCATGAGTTTGACAGCTGAGTTCACAGGCGCTT 224
QY 61 AlAValAlAArgGgIyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIle 80
DB 225 AGTGCCTCACTGAGAGTTCCTCTCAG----- 251
QY 81 TyrValIleThrGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
DB 251 ----- 251
QY 101 TyrGlyAlaProProAlaGlyTyrGlyAlaAlaProProGlyTyrGlyAlaProProAla 120
DB 252 -----CCCCCTGCTATGATACACACCTGTA 278
QY 121 GlyTyrGlyAlaGlnProValGlyAsnGluGlyProProVal-----GlyTyrArgAla 138
DB 279 CCTGAGAGGTATGACCTGTC-----CCAGCTGCTCAGAGAGGTATGACCT 326
QY 139 SerProValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGly 158
DB 327 GCTCCAGAGAGGTATGACCTGCTCAGAGAGGTATGACCTGCTCAGAGAGGTATGCA 386
QY 159 AlaProProLeuGlyTyrGlyAlaPro-----ProLeuGlyTyrGly 172
DB 387 CCTGCTCCGAGGAGGTATGCTCTCTCCACACCTCCAAATGCTTATCTTATGCA 446
QY 173 ThrProProLeu-----GlyTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGly 191
DB 447 CCACCTCCGATGAATGCTTACGAGACGTCTCCAGCCCATGGGTATCTGTATGCCAG 506
QY 192 AsnGluGly-----ProPro 196
DB 507 AATCAGGTATTATACCAACCG 527
RESULT 11
LOCUS BU107273 932 bp mRNA linear EST 25-NOV-2002
DEFINITION 603111954F1 CSBQCHL12 Gallus gallus cDNA clone CHEST60k13 5', mRNA
SEQUENCE
ACCESSION BU107273
VERSION BU107273.1 GI:25309754
KEYWORDS EST
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Phasianidae; Gallus.
REFERENCE Boardman, P.B., Sanz-Bzquez, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.,
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
2235534
JOURNAL PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 10D, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES
Source location/Qualifiers
1..932
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST60K13"
/dev_stage="36"
/lab_host="DH10B"
/clone_1ib="CSBQCHL12"

/note="Organ: heart; Vector: pBluescript II KS(+); Site: 1:
 EcoRI; Site: 2: NotI; Modification of pBluescript II KS(+)
 [Stratagene] vector to accommodate cDNA produced with the
 T-primed protocol (Construction of uni-directionally
 cloned cDNA libraries from messenger RNA for improved 3'
 end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,187
 ,644). Cut pBluescript II KS(+) with NotI and EcoRI.
 Ligate in double stranded adaptor containing BspI and
 BamHI sites (5'-ggcggcgccagcccgagcgaataaag-
 3' aattcttttcgagtcgagcgagcgagcgcg-5')

BASE COUNT 215 a 273 c 233 g 211 t
 ORIGIN

Alignment Scores:

Pred. No.: 3,19e-13 Length: 932
 Score: 338.00 Matches: 88
 Percent Similarity: 42.56% Conservative: 15
 Best Local Similarity: 36.36% Mismatches: 64
 Query Match: 26.43% Indels: 76
 DB: 13 Gaps: 7

US-09-864-291-12 (1-235) x BU107273 (1-932)

Qy 1 MetProPheAplLeuMetThrAsnLeuThrValGlnInProValPheAlaAlaAspHe 20
 Db 231 ATGCCGTTTATTGTAAGAAAGGCTCTATTGAGAGAGCTCTTCTCTGCTAATTAC 230
 Qy 21 IleYsGlyThrIleGlnAlaAlaProTyrglyYTrpGlnGlyGlnAlaThrPheYs 40
 Db 291 ATCAAAGGACAGATTCAAGCTGAGAGGAGAGGTGGTGGAGAGGAGGAGAAAGTTTAA 350
 Qy 41 LeuValPheArgAsnGlyAspAlaIleGlnPheAlaGlnLeuMetValYsAlaAlaSer 60
 Db 351 CTACCTTTCAACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 410
 Qy 61 AlaValAlaArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIle 80
 Db 411 AGTGCTTCAGTGAAGTCTCTCCAG- 437
 Qy 81 TyrValIleThrGlyGlnGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
 Db 437 - 437
 Qy 101 TyrGlyAlaProProAlaGlyTyrglyAlaProProGlyTyrglyAlaProProAla 120
 Db 438 - CCCCCTGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 464
 Qy 121 GlyTyrglyAlaGlnProValGlyAsnGlnGlyProProVal- -GlyTyrgAla 138
 Db 465 CTGGAGGGGATGACCTGTC- -CCACTGCTCCAGAGGGGATGACCT 512
 Qy 139 SerProValArgTyrglyAlaProProLeuGlyTyrglyAlaProProAlaGlyTyrgly 158
 Db 513 GCTCCAGGAGGATGAGCTGCTCCAGAGGGGATGACCTGCTCCAGAGGGGATGAC 572
 Qy 159 AlaProProLeuGlyTyrglyAlaProPro- -LeuGlyTyrgly 172
 Db 573 CCTGCTCCGGGGGCTATGCTCTCTCCAGACCTCCAAATGCTTATCC- -TATGCA 631
 Qy 173 ThrProProLeu- -GlyTyrglyAlaProProLeuGlyTyrglyAlaProProAlaGly 191
 Db 632 CCACCTCCAGTAATGCTTACGAGCTGCTCCAGCCCATGGGATATCCGATGCCCCAG 651
 Qy 192 AsnGlnGly- -ProProAla 197
 Db 692 AATCCAGGATATTAACCCACGCTCCACATGATGATGATGATGATGATGATGATGATG 751
 Qy 198 GlyTyrgAlaSerProAlaGlySerGlyAlaArgProGlnGlnGlnAlaAlaGln 217
 Db 752 CCCTAC- -TCCGGGCACTCTCTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 793
 Qy 218 AlaPro 219

Db 794 CTTCCA 799

RESULT 12

CD301781

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CDNA LIBRARY PREPARATION: Life Technologies, Inc.

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLNL4587 row: a column: 21

High quality sequence stop: 646.

Location/Qualifiers

1..892

/organism="Xenopus laevis"

/mol_type="mRNA"

/db_xref="taxon:8355"

/clone="IMAGE:6957790"

/dev_stage="adult"

/lab_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: PCMV-SpOx; Site: 1: NotI;

Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 2.3 kb. Constructed by Life

Technologies. Note: This is a Xenopus Gene Collection (XGC

) library."

BASE COUNT 239 a 247 c 198 g 208 t

ORIGIN

Alignment Scores:

Pred. No.: 2.77e-12 Length: 892

Score: 333.50 Matches: 91

Percent Similarity: 45.27% Conservative: 19

Best Local Similarity: 37.45% Mismatches: 74

Query Match: 25.29% Indels: 59

DB: 14 Gaps: 11

US-09-864-291-12 (1-235) x CD301781 (1-892)

Qy 1 MetProPheAplLeuMetThrAsnLeuThrValGlnInProValPheAlaAlaAspHe 20

Db 254 ATGCCGTTTATTGTAAGAAAGGCTCTATTGAGAGAGCTCTTCTCTGCTAATTAC 313

Qy 21 IleYsGlyThrIleGlnAlaAlaProTyrglyYTrpGlnGlyGlnAlaThrPheYs 40

Db 314 ATCAAAGGACAGATTCAAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 373

Qy 41 LeuValPheArgAsnGlyAspAlaIleGlnPheAlaGlnLeuMetValYsAlaAlaSer 60

Db 374 CTTACATTAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 433

Qy 61 AlaValAlaArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIle 80

ACCESSION CD255258
 VERSION CD255258.1 GI:31015724
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM Xenopus laevis
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae; Xenopodinae; Xenopus.
 REFERENCE 1 (bases 1 to 868)
 AUTHORS NIH-WGC <http://wgc.nci.nih.gov/>.
 TITL National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabs-remail.nih.gov
 Tissue Procurement:
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LML at: <http://image.llnl.gov>
 Plate: LLM14578 row: 9 column: 04
 High quality sequence stop: 673.
 Location/Qualifiers
 1..868
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:5954461"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /clone_1lb="NICHD_XGC Bml4"
 /note="Organ: brain; Vector: PCMV-SPORE; Site 1: NCI; Site 2: Sali; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.5 kb. Constructed by Life Technologies. Note: This is a Xenopus Gene Collection (XGC) library."
 BASE COUNT 241 a 244 c 189 g 193 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 6.24e-12 Length: 868
 Score: 318.00 Matches: 85
 Percent Similarity: 44.74% Conservative: 17
 Best Local Similarity: 37.28% Mismatches: 78
 Query Match: 24.86% Indels: 48
 DB: 14 Gaps: 8
 US-09-864-291-12 (1-235) x CD255258 (1-868)
 QY 1 MePProPheAepLeuMetThrAsnLeuThValGluGlnProValPheAlaAlaAsnPh 20
 DB 253 ATGCAATTCCTACCTGATGAAGAGATGCTCCATTAACACGCACTTCTCTGCAACTAC 312
 QY 21 ILeYsgLYThrIleGlnAlaAlaProTYrGlyTYrGlyGlnAlaThrPheLYs 40
 DB 313 ATCAAGGGAACCATCAGTCTGAACCTGAGAGCTGAGAGGCTAAACCTCTTTAAA 372
 QY 41 LeuValPheArGaAnGlyAspAlaIleGluPheAlaGlnLeuMetValLYsAlaAlaSer 60
 DB 373 CTTCATTTCACACTGAGAGAGCTATGAAATTTGACAAATCATGTTCAAAATGCAACT 432
 QY 61 AlaValAlaArGlyLYpPheProLeuArGThrLeuAenAPrTPheSerSerMetGlyIle 80
 DB 433 TCTGCTTCAGAGCCCTCTCTGT----- 456
 QY 81 TyrValIleThrGlyGlyGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
 DB 456 ----- 456

QY 101 TYrGlyAlaProProAlaGlyTYrGlyAlaProProGlyTYrGlyAlaProProAla 120
 DB 457 -----CTTCAGCTGCTCATATGATATCACTCTGCTGAGAGATATGACAGGC 507
 QY 121 GLYTYrGlyAlaGlnProValIleAsnGlyLYpProProValGlyTYrArGAlaSerPro 140
 DB 508 GGGTACCCCTCCAGCTCCAGGAATATATACACCCCAACC-----CAACCATCAGGCCCT 561
 QY 141 ValArGTYrGlyAlaProProLeu---GlyTYrGlyAlaProProAlaGlyTYrGly--- 158
 DB 562 TATTCATATGACCCCTGCTATGATGATATGACACCTCCCAATGCTATGAGATAC 621
 QY 159 -----AlaProProLeuGlyTYrGlyAlaProProLeuGlyTYrGlyThrProProLeu 176
 DB 622 CCGTATGCTCCGACACAGACCTGCAATATCAACCCCTCCGATATGATATCCATT 681
 QY 177 GLYTYrGlyAlaProProLeuGlyTYrGlyAlaProProAlaGlyAsnGlyLYpProPro 196
 DB 682 ---TACATGCTCTCCACACCTTATCTTGACCCCG-----TATATGACACCCCA 732
 QY 197 AlaGlyTYrArGAlaSerProAlaGlySerGlyAlaArGProGlnGlnSerThrAlaAla 216
 DB 733 ---TATATGAAACCCCAAGACCTTACGCT-----CCCTGATTCATGATGACT 780
 QY 217 GlnAlaProGlnAenGlnAlaSer 224
 DB 781 GGTATGCTGAGAGGACGACCAAGC 804
 RESULT 15
 BQ735296 862 bp mRNA linear EST 16-JUL-2002
 LOCUS AGENCOURT 8096274 NICHD_XGC Bml4 Xenopus laevis cDNA clone
 DEFINITION IMAGE:5543298 5', mRNA sequence.
 BQ735296
 BQ735296.1 GI:21874193
 EST.
 ACCESSION Xenopus laevis (African clawed frog)
 VERSION Xenopus laevis
 KEYWORDS BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae; Xenopodinae; Xenopus.
 SOURCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 ORGANISM National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Igor David
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LML at: www-bio.llnl.gov/bdrrp/image/html
 Plate: LLM12243 row: c column: 21
 High quality sequence stop: 710.
 Location/Qualifiers
 1..862
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:5543298"
 /dev_stage="embryo, stage 31-32"
 /lab_host="DH10B (phage-resistant)"
 /clone_1lb="NICHD_XGC Bml4"
 /note="Organ: whole embryo; Vector: PCMV-SPORE; Site 1: NCI; Site 2: Sali; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 kb. Constructed by Life Technologies. Note: This is a Xenopus Gene Collection (XGC) library."
 BASE COUNT 233 a 247 c 186 g 195 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.:	1.54e-11	Length:	862
Score:	312.00	Matches:	82
Percent Similarity:	43.04%	Conservative:	17
Best Local Similarity:	35.65%	Mismatches:	71
Query Match:	24.39%	Indels:	60
DB:	13	Gaps:	6

US-09-864-291-12 (1-235) x BQ735296 (1-862)

```
OY      1 MetProPheAepLeuMetThrAsnLeuThrValGluGlnProValPheAlaAlaSerPhe 20
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      234 ATGCATTCTACTGATGAAAGAGTCTCATAGACACCGCTCTTCTGCAACTAC 293
OY      21 ILeYsgLYThrIleGlnAlaAProTYrgLYTYrgLYGlnAlaThrPheLYs 40
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      294 ATCAAGGAGACATCATGCTGACCTGAGATGGCTGGAGGTAACCTTTTAA 353
OY      41 LeuValPheArgAsnGlyAspAlaIleGluPheAlaGlnLeuMetValIleAlaSer 60
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      354 CTTACATTAAACAGTGAAGAGACTATTGAATTGGAACAATCATGTTCAAAATGGCAACT 413
OY      61 AlaValAlaArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIle 80
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      414 TCTGCTTCAGAGCCCTCTCTGTT----- 437
OY      81 TYrValIleThrGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      437 ----- 437
OY      101 TYrgLYAlaProProAlaGlyTYrgLYAlaProProGlyTYrgLYAlaProProAla 120
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      438 -----CCTCATGCTGATATGATATGACTCTGCTGACAGAGATATGACACAGGC 488
OY      121 GlyTYrgLYAlaGlnProValGlyAsnGluGlyProProValGlyTYrArgAlaSerPro 140
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      489 GGGTACCTCCAGCTCCAGGAATATATACACCCCAACC-----CAACCATCAGGCCCT 542
OY      141 ValArgTYrgLYAlaProProLeu---GlyTYrgLYAlaProProAlaGlyTYrgLY--- 158
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      543 TATCATATGACACCCCTGCTATGATATGACATGACACCTCCGAATGCTATGGGATAC 602
OY      159 -----AlaProProLeuGlyTYrgLYAlaProProLeuGlyTYrgLYThrProProLeu 176
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      603 CCGTATGCTCCGCGCACAGACCTGGCATGTATCCACCCCTCTGAATGATCCCATT 662
OY      177 GlyTYrgLYAlaProProLeuGlyTYrgLYAlaProProAla----- 190
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      663 ---TACATGGCTCTCCACACCACTTATCTGGACCCCGGTAAATGACCCCATATAA 719
OY      191 -----GlyAsnGluGlyProProAlaGly 198
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      720 TGGAAACCCGACGACCTTACGCTCCCTGAGATCCATGGATATCTGGGTATGCTTGAAGGC 779
OY      199 TYrArgAlaSerProAlaGlySerGlyAla 208
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      780 AGCAAGCCACAGAAAGCAGCATCCAGTGCA 809
```

Search completed: December 16, 2003, 20:45:15
Job time : 1526.57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 05:59:34 / Search time 5255.22 seconds

(without alignments)
10999.596 Million cell updates/sec

Title: US-09-864-291-4

Sequence: 1 cggcaccgagggcgcgagga.....gcataaaaaaaaaaaaaa 1413

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

```
Database :
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_by:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_ey:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1413	100.0	1413	6	AX359658
2	592.4	41.9	2267	9	BC022546
3	587.2	41.6	2266	9	BC022549
4	435.4	30.8	179222	6	AX359665
5	347.2	24.6	179222	2	BX296515
6	347.2	24.6	223469	2	BX470149
7	282.6	20.0	220895	9	H8250D10
8	223.2	15.8	129624	2	BX296540
9	188.8	13.4	168425	2	AC113593
10	188.8	13.4	224086	10	AC104325
11	172.6	12.2	253149	2	AC107527
12	172.6	12.2	270171	2	AC133969
13	128.4	9.1	793	10	AF499026
14	128.4	9.1	1802	10	MF040826
15	122.8	8.7	223469	2	BX470149
16	122	8.6	1691	9	HSU79458
17	122	8.6	1820	9	BC007452
18	122	8.6	1885	6	AR220846
19	122	8.6	1912	9	BC010616
20	121.6	8.6	2170	9	AB072784
21	120.4	8.5	1835	9	AK057881
22	116	8.2	1205	5	BC053332
23	116	8.2	95108	2	AC127818
24	110.2	7.8	106438	9	AL365445
25	106.4	7.5	2065	14	AF410847
26	106.4	7.5	148418	9	AC020698
27	103.8	7.3	454	6	BD027759
28	102.8	7.3	246814	2	AC103482
29	100	7.1	235381	2	AC103031
30	96	6.8	247947	2	AC098163
31	95.8	6.8	110000	2	AC118467_0
32	93.8	6.6	216292	2	AC100927
33	91.6	6.5	239972	2	AC095835
34	91.6	6.5	241519	2	AC097689
35	91.6	6.5	250902	2	AC131965
36	91.6	6.5	251746	2	AC131411
37	91.4	6.5	208880	2	AC128103
38	90.6	6.4	230569	2	AC108250
39	90.6	6.4	238377	2	AC121702
40	90.4	6.4	210680	2	AC123724
41	90.4	6.4	213334	2	AC023070
42	90.4	6.4	315390	2	AC124097
43	90	6.4	262967	2	AC110392
44	89.8	6.4	261914	2	AC122581
45	88.8	6.3	88716	2	AC098644

ALIGNMENTS

RESULT 1
LOCUS AX359658 1413 bp DNA linear PAT 13-FEB-2002
DEFINITION Sequence 4 from Patent WO0190185.
ACCESSION AX359658
VERSION AX359658.1 GI:18675409
KEYWORDS
SOURCE unidentified
ORGANISM unclassified
REFERENCES
1
Oko, R. and Sutovsky, P.
Pl32 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses thereof
JOURNAL Patent: WO 0190185-A 4 29-NOV-2001;

QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES
UNIVERSITY (US)
Location/Qualifiers
1. 1413
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="mammalia"
36.977
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD3336.1"
/db_xref="GI:18675410"
/translation="MAVNSHTSRBRLALPSGSVULKQEDVDLCTLOKRVSYLEN
GTRKGLFLTSYRVFTSHLVNDPMLSFMPPLMSDCTTBPTAPNTIKTIDAA
PQGGMEQAVFKLSFRKGAIBPAQLWVKASAAARIPASVNYWFDGLYITVP
GAAVCSQTECPAIPVIGPBPBGATVQGBYTPREGVDAQOGKGGAPMGVADP
VGVGPBGYGVPPGPGYGAAPPGVGPBGYGAAPPGYGAAPPGYGAAPPGYGAAP
NEALPPEYEAHPAGNTAASHRSMTAQOETSLPTSSS"

CDS

BASE COUNT 377 a 363 c 369 g 304 t
ORIGIN

Query Match 100.0%; Score 1413; DB 6; Length 1413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGACAGAGGGCGCGAGAGGGGCGCTGGGCAAGATGACATGACACGACCG 60
DB 1 CGGACAGAGGGCGCGAGAGGGGCGCTGGGCAAGATGACATGACACGACCG 60
QY AGAGCCGTCGAGGGGCGCTGATCCCTGCGGCAAGATGCTTGAAGCAAGTGAAGATG 120
DB AGAGCCGTCGAGGGGCGCTGATCCCTGCGGCAAGATGCTTGAAGCAAGTGAAGATG 120
QY 121 TGAACCTCTGCTCTCTCAAGAAACAGTGAATCTTCTTTAATGCAAAAGAAAG 180
DB 121 TGAACCTCTGCTCTCTCAAGAAACAGTGAATCTTCTTTAATGCAAAAGAAAG 180
QY 181 GAAAGCTGTTTCTCACTTACATACCGGAGTGTCTTGATGACTTCACTTATGATGACC 240
DB 181 GAAAGCTGTTTCTCACTTACATACCGGAGTGTCTTGATGACTTCACTTATGATGACC 240
QY 241 CGATGCTTCTTTAATGATGCGCTTGGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 300
DB 241 CGATGCTTCTTTAATGATGCGCTTGGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 300
QY 301 TTTTGGCCCCCACTCAATTAAGAAACATTTCAAGGAGCTCCAGTGTGGTGGAAAG 360
DB 301 TTTTGGCCCCCACTCAATTAAGAAACATTTCAAGGAGCTCCAGTGTGGTGGAAAG 360
QY 361 GACAAAGCTGTTTAAAGTATCTTCAAGAAAGAGTGCATGCAATTTGCCCACTGA 420
DB 361 GACAAAGCTGTTTAAAGTATCTTCAAGAAAGAGTGCATGCAATTTGCCCACTGA 420
QY 421 TGGTAAAGCTGCTCTGCTGCTGCAAGAAATTCCTTGAAGTGAATTAATCTGGT 480
DB 421 TGGTAAAGCTGCTCTGCTGCTGCAAGAAATTCCTTGAAGTGAATTAATCTGGT 480
QY 481 TCGACACTTCAAGACCTGATCAATTAATTAATCTCCAGGGGCTGAGTGTCTCTCAAGA 540
DB 481 TCGACACTTCAAGACCTGATCAATTAATTAATCTCCAGGGGCTGAGTGTCTCTCAAGA 540
QY 541 CACCTTGTCCAGCATATCCAAATTGTGATCTATGAGCCCAACACCAAGATTAACATGCC 600
DB 541 CACCTTGTCCAGCATATCCAAATTGTGATCTATGAGCCCAACACCAAGATTAACATGCC 600
QY 601 AACCGGGGAATATGAACTCCACCAAGAAATATGAGGCCCAACCAAGGGGAATATGAG 660
DB 601 AACCGGGGAATATGAACTCCACCAAGAAATATGAGGCCCAACCAAGGGGAATATGAG 660
QY 661 CCCGACCTATGAGATATGAGCCCGGCTGTGGATATGAGATCCCACTGGGGGAATATG 720
DB 661 CCCGACCTATGAGATATGAGCCCGGCTGTGGATATGAGATCCCACTGGGGGAATATG 720

QY 721 GAGTCCCACTGGGGGAATATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAAT 780
DB 721 GAGTCCCACTGGGGGAATATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAAT 780
QY 781 ATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAAT 840
DB 781 ATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAAT 840
QY 841 GATATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAAT 900
DB 841 GATATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAAT 900
QY 901 CTGAAATATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAAT 960
DB 901 CTGAAATATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAATATGAGATCCCACTGGGGGAAT 960
QY 961 CTACCTCATCTTCTTATGATGATCAATTTAACAATCTTCAAGATTAACCTTGAAGATCACTCAC 1020
DB 961 CTACCTCATCTTCTTATGATGATCAATTTAACAATCTTCAAGATTAACCTTGAAGATCACTCAC 1020
QY 1021 AAGCAAAAGGCAACCTTAAATGAACTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1080
DB 1021 AAGCAAAAGGCAACCTTAAATGAACTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1080
QY 1081 TAGAAGATGTCGTGTGACAGCAAGTGTGATCTTCTTCAACAACCTTGAAGTGTCTGT 1140
DB 1081 TAGAAGATGTCGTGTGACAGCAAGTGTGATCTTCTTCAACAACCTTGAAGTGTCTGT 1140
QY 1141 GCTCAAAACAGATGAAGGTGAGAAAGCACTCTCTTCTCAAGAAAGAAAGTCTTGA 1200
DB 1141 GCTCAAAACAGATGAAGGTGAGAAAGCACTCTCTTCTCAAGAAAGAAAGTCTTGA 1200
QY 1201 AAACAGATGCAAGCCCACTGAAGAGAGAGAGATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1260
DB 1201 AAACAGATGCAAGCCCACTGAAGAGAGAGAGATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1260
QY 1261 GAATGAGATGATCTCTTGAAGAAACAGGCTTCTCTCTGCTGCTGATGATGATGATGATGATGATGATGAT 1320
DB 1261 GAATGAGATGATCTCTTGAAGAAACAGGCTTCTCTCTGCTGCTGATGATGATGATGATGATGATGATGAT 1320
QY 1321 AAGAAATGCTGAGAACCAAGAGTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1380
DB 1321 AAGAAATGCTGAGAACCAAGAGTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1380
QY 1381 CTTGCTTCTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1443
DB 1381 CTTGCTTCTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1443

RESULT 2
BC022546 2267 bp mRNA linear PRI 04-FEB-2002
LOCUS
DEFINITION
Homo sapiens, similar to RIKEN cDNA 4930521J3 gene, clone
MGC:26816 IMAGE:4811804, mRNA, complete cds.
BC022546
BC022546.1 GI:18490706
MGC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2267)
Strauberg, R.
Direct Submission
Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

REMARK
COMMENT

CDNA library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
DNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shngc.stanford.edu>
Contact: (Dickson, Mark) mcdpaxaki@stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINAC at: <http://image.llnl.gov>
 Series: IRAC Plate: 32 Row: 3 Column: 20
 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES

Source

```

source
1. .2267
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26816 IMAGE:4811804"
/class="type="Brain", hippocampus"
/clone_lib="NIH MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
32. .961
DSDS

```

CDS

```

//sequence="similar to RIKEN CDNA 493052173 gene"
//product="similar to RIKEN CDNA 493052173 gene"
//protein_id="AAH2546.1"
/db_xref="GI:18490707"
/translation="MAVNSQSTHTNRGALIPNGESILKRSFNVELSPFORSGNSVFS
GRKQTJLFLTSYRVIFPTSCSIDPMLSPMPPLDMLTLTYEQVYPAAPTKGTQAP
PYGMEGQATKFLVFRNGGALIFQMLTKAASLAARGPLTLNDMFSSMGIYYTGG
GNMCTPQMFCSVIYYGAPAGYGAPBPSPYGAPQVGNESBPVGYASPRVYGY
APLPYGAPAGYGYGAPPLGYGAPPLGYCTPLPGYCAPPLGYAPAGNIGBPAGYRAS
PAGGAPRHESTAAQAPNEASLSPSASSQVHS"

```

Query Match	41.9%	Score 592.4;	DB 9;	Length 2267;
Best Local Similarity	75.7%	Pred. No. 5.9e-136;		
Matches 796;	Conservative 0;	Mismatches 231;	Indels 25;	Gaps 4;

Qy	21	GGGGGCTCGGCAAGATGGCAGTAAACAAGACCAACCGAAGCCGTGTGGGGCTTC	80
Db	17	GAGGCCCGAACAAGATGGCGGTGAATCAGAGCCAACAAGAAACCGCGGAGCCCTTC	76
Qy	81	ATCCCTCTGCGAAGAGTCTTTGAAGCAGTGAAGATGGACCTCTTGCTTCCTAAG	140
Db	77	ATCCCTAACGTGAAAGTCTTTGAAGGGGTCTCGAATGTGAGACTCTCTCCCAAG	136
Qy	141	AAACCAAGGAATCCTATCTCTTTAAATGGCAAAAGAAAGAAAGTGTGTCTCACTTCA	200
Db	137	CGATCAGAAAGGCTCAAAATGTCTTTAAGTGTGAAGAAAGACAGAAACATGTCTCACTTCA	196
Qy	201	TACGGGGGTGCTTCGTGACTTCAACTAGTCAATGAACCCCATGCTTCTTTATGATG	260
Db	197	TACGGGGTATTTTCATTACTTCATGCTTCATCACTGATATCCCATGTGTCTTTATGATG	256
Qy	261	CCGTTTGGCCTGATGATGATGACATGCACATTTGAACAACCAATTTTGGCCCCCACTACAT	320
Db	257	CCATTTGATCTGAATGACGAACTCACTGTTHAACACAGATATTTGCTGCAAACTTCAAT	316
Qy	321	AAAGAACATTCAGGCACTCCAGGTGTGTGGCTGGGAAGACAAGCTGTATTTAAAGTTA	380
Db	317	AAAGGAATATTCAGGCACTCCATATGTGTGGGAAGACAAGCTACTTTTAAATTA	376
Qy	381	TCCTTCAGGAAGAAGGTGCATCGAATTTGGCCAACTGATGTAAAAGTGCCTGTGCT	440
Db	377	GTCCTTCAGGAATGAGAGTGCATTTGAATTTGGCCAGTTGATGTAAAGTGCCTGTGCT	436
Qy	441	GCTGCGAAGAAATTCACCTTGAAGTGAATTAATGTGGTTGCACCTTCAGAGCTGAC	500

Db 437 GCTGCCGAGGATTTCACCTTAGAAGCTTAATGACTGGTTCAAGCTCTAATGGAAATTAT 496

Qy 501 ATAAATTACTGTCCAGAGGGCTGCAGTGTGCTCTCAACAGACACTTGTCCAGCATATCCA 560

Db 497 GTAATTACTGGGAGAGGAAT---ATGTGACCTCCACAGATGCTTGTTCAG----- 545

Qy 561 ATTGTGATCTATGAGACCCCAACCAAGATATATCACTGTCGAACCAAGGGAAATATGAACT 620

Db 546 -TTATGTGCTATGAGAGCCCACTGTCAGAGATATAGAGGCCCACTCCCGATACGAGGCC 604

Qy 621 CCACCAAGAGATATGAGGCCCAACGAGGGGATATAGAGCCCACTATGGATATGGA 680

Db 605 CCACCTCAGAGATATGAGGCCCAACCGTATGGAATATAGAGCCCGCTGTGAGATACAGA 664

Qy 681 GCCCGGCTGTGGGATATGAGAGTCCCACTGAGGGGAAATATGAGAGCCCACTGGGGGATAT 740

Db 665 GCTCACTCTGTGCAATATGAGGCCCACTCTTGGATATGAGAGCCCACTTGCAAGATAT 724

Qy 741 GGAAGTCCCACTGGGGGATATGAGGCCCACTGGGGGATATGAGTCCCACTGGGGGA 800

Db 725 GGAAGCCCACTCTAGAGATATGAGGCCCACTTATGGAATATGGAACCCCACTCTGGA 784

Qy 801 TATGTGAGCCCACTTGGGGGATATATGAGGCCCACTGCAAGATATGAGAGCCCACTAGCT 860

Db 785 TATGAGAGCCCACTCTGAGATATGAGGCCCACTGCAAGAGAAATGAGAGCCCGCTCGCG 844

Qy 861 GGAATGAGAGCCCACTGAGATATGAGAGCTCACTGTGGAATATAGAGTGCCTCT 920

Db 845 GGAATACAGAGCTCACTGCTGTGATACAGAGCCAGGCTCAAGATATTACAGAGCC--- 901

Qy 921 CACAGATCTATGACAGCTCAGCAGAGAGACTTCTCTTCCCACTACCTCATCTTCTTAAAGTC 980

Db 902 -----CAGGCTCTGAGAAAGAGAGCTTCTTCTCTGCTGCTCTTCTTCAGAGTC 952

Qy 981 CATTACCACTTCTCAGAGTTAAACCTTGAAGCTCAGCAAGAAAG--GGACCCATAA 1039

Db 953 CATTCTTAACCTTCAAGATGTAAACCTTGAAGCTCAACAGCAAGAGGTACCTTAA 1012

Qy 1040 ACTGAAGTCACAGTAAGAGGAAAGCCAGGT 1071

Db 1013 ATTGAAGTCAGATATAGAGAGCAACTCAGGT 1044

[illegible]REMARK
COMMENT

Email: CSapals-rc@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
DNA Library Preparation: Michael J. Brownstein (NHGRI) & Toshiyuki and Piero Carninci (RIKEN)
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-nhgsc.stanford.edu>

Contact: (Dickson, Mark) mcd@pax1.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
 Series: IRAX Plate: 32 Row: m Column: 3
 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES
 source
 Location/Qualifiers
 1..2266
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:26828 IMAGE:4815949"
 /issue_type="Brain, hippocampus"
 /clone_lib="NH MGC_95"
 /lab_host="DH10B"
 /note="Vector: pBlueScript"

CDS

32..634
 /codon_start=1
 /product="Similar to RIKEN cDNA 4930521123 gene"
 /protein_id="AAH2549.1"
 /db_xref="GI:18490711"
 /translation="MAVNSHTENRNGALIPNGESILKRSFNVLSFPQRBGSNVPS
 GRKTLPLVSTRVLTPTSCSIDPMLSPMPDPDLMTNLTVBQVPVAPFVIGTQIA
 PYGEGQATFLVFNQGAIBFQALMVPAASAAAPFLARTLNMFSMGIVYVIGK
 GICALHRCVQLSMEPHLQDMEPHLDPDEPHLQDMEPHP"

BASE COUNT 709 a 477 c 468 g 612 t
 ORIGIN

Query Match 41.6%; Score 587.2; DB 9; Length 2266;
 Best Local Similarity 75.4%; Pred. No. 1.2e-134;
 Matches 793; Conservative 0; Mismatches 233; Indels 26; Gaps 4;

```

QY 21 GGGGGCGCTGGGAGATGCGAGTGAACGACGACCAACGAGAGCGCTGCTGGGCGCCCTC 80
DB 17 GAGGCGCCGAGAGAGATGCGGATGTAATCAGAGCCACGAGAAACCGCGAGGCGCTC 76
QY 81 ATCCCTCTGGCGAAAGTGTCTTGAAGAGTGAAGATGAGACCTGCTTCTTACAG 140
DB 77 ATCCCTAAGCGTGAAGTGTCTTGAAGCGGTCTCCGAATGTGAGCTCTCTCCACAG 136
QY 141 AAACGAGTGAATCTATCTCTTTAATGCGACAAAGAAAGAAAGTGTCTTCACTTCA 200
DB 137 CCAATCAGAGGCTCAATATGCTTAGTGTGAAGAAAGACAGAAATGTTTCTCACTTCA 196
QY 201 TACCGGCTGTCTCTGTAATCTTCACTTGAATGACCCCAATGCTTCTTTAATGAG 260
DB 197 TACCGGCTGTATTTCAATCTTCAATCTCCTCAGTGAATCCCAATGTTGCTTTAATGAG 256
QY 261 CGGTATGCTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 320
DB 257 CCAATTTGATCTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 316
QY 321 AAAGAAACCAATTCAGGAGCTCCAGGTGTGCTGAGAAAGCAAGCTGTTTAAAGTGA 380
DB 317 AAGGAAACCAATTCAGGAGCTCCAGGTGTGCTGAGAAAGCAAGCTTAAATTA 376
QY 381 TCCCTTGAAGAAAGAGTGTGCTGATGATTTGCCCACTGATGTAAAGTGTGCTGCT 440
DB 377 GTCTTCAAGAAATGAGAGTGTGCTGATGATTTGCCCACTGATGTAAAGTGTGCTGCT 436
QY 441 GCGCGAGAGGAATTCAGTGTGAAGTGAATTAATGATGTGAAGTGAAGTGAAGTGAAGT 500
DB 437 GCGCGAGAGGAATTCAGTGTGAAGTGAATTAATGATGTGAAGTGAAGTGAAGTGAAGT 496
QY 501 ATTAATTAATGCTGCGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 560
DB 497 GTAAATTAATGCTGCGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 544
QY 561 ATTGTGATCTAATGAGACCCCAACACAGATATATACAGTCCAAACAGAGGAAATATGAACT 620

```

```

DB 545 -TTATGCTATGAGAGCCCACTGACAGATATGAGAGCCCACTCCCGAATACGAGCC 603
QY 621 CCACGAGAGATATGAGAGCCCACTGACAGATATGAGAGCCCACTATGAGATATGAG 680
DB 604 CCACCTGAGAGATATGAGAGCCCACTGACAGATATGAGAGCCCACTGTGAGATACAG 663
QY 681 GCCCGCGCTGAGAGATATGAGAGCCCACTGACAGATATGAGAGCCCACTGTGAGATAT 740
DB 664 GCTTCACTGAGAGATATGAGAGCCCACTGACAGATATGAGAGCCCACTGTGAGATAT 723
QY 741 GAGTCCCACTGAGAGATATGAGAGCCCACTGACAGATATGAGAGCCCACTGTGAGATAT 800
DB 724 GAGGCGCCCACTGAGATATGAGAGCCCACTGACAGATATGAGAGCCCACTGTGAGATAT 783
QY 801 TATGAGTCCCACTGAGAGATATGAGAGCCCACTGACAGATATGAGAGCCCACTGTGAGATAT 860
DB 784 TATGAGAGCCCACTGAGATATGAGAGCCCACTGACAGATATGAGAGCCCACTGTGAGATAT 843
QY 861 GGAATATGAGAGCCCACTGAGATATGAGAGCCCACTGACAGATATGAGAGCCCACTGTGAGATAT 920
DB 844 GATATCAGAGCCCACTGAGATATGAGAGCCCACTGACAGATATGAGAGCCCACTGTGAGATAT 900
QY 921 CACAGATCTATGACAGCTCAGAGAGAGACTTCTTCCCACTCACTCACTTCTTAGGTC 980
DB 901 -----CAGGCTCTGAGAAAGAGAGCTTCTTCCCTGCTGCTCTTCTTCTGAGTC 951
QY 981 CATTTACCACTTCTCAGAGTAAACCTTGAAGACTCAACAGAGAG-AGCACTTAA 1039
DB 952 CATCTTAACTTCTTAAATATGATTAACCTTGAAGACTCAACAGAGAGAGGATACCTTAA 1011
QY 1040 ACTGAAGTCAAGTAAAGAGAGAGAGAGT 1071
DB 1012 ATTTAAGTCAAGATTAAGAGAGAGAGAGT 1043

```

RESULT 4
 AX359665
 LOCUS AX359665 1001 bp DNA linear PAT 13-FEB-2002
 DEFINITION Sequence 11 from Patent WO0190185.
 ACCESSION AX359665
 VERSION AX359665.1 GI:18675411
 KEYWORDS

SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Buxaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1. Oko, R. and Sutovsky, P.
 Pt3 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses
 thereof
 Patent: WO 0190185-A 11 29-NOV-2001;
 QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES
 UNIVERSITY (US)

JOURNAL
 TITLE
 AUTHORS
 JOURNAL

FEATURES
 source
 Location/Qualifiers
 1..1001
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 1..708
 /note="unnamed protein product"

CDS
 /codon_start=1
 /protein_id="CAD23337.1"
 /db_xref="GI:18675411"
 /translation="MPPDLMTNLTVBQVPVAPFVIGTQIAAPYGEAGTATFLVFR
 NGDAIRPAQIVKASAVARGPPLRTDMSGIVYVIRGAGACGTCMGCSVYVG
 APPAGYCAPPGYCAPPGYCAPPGYCAPPGYCAPPGYCAPPGYCAPPGYCAPPGYCAPPGYCAP
 PLGAGAPPLGAGPPLGAGPPLGAGPPLGAGPPLGAGPPLGAGPPLGAGPPLGAGPPLGAGPPLGAGP
 PENASLPSASBSOVHS"

BASE COUNT 261 a 254 c 248 g 238 t
 ORIGIN

Query Match 30.8%; Score 435.4; DB 6; Length 1001;

Best Local Similarity 74.78; Pred. No. 5.3e-97;
Matches 609; Conservative 0; Mismatches 181; Indels 25; Gaps 4;

QY 258 ATGCGGTTGGCCGATGAGTACGACCAATGAAACCAATTTTCCCAACAC 317
|||||
Db 1 ATGCGATTGATGATGACCACTCACTGTGAACCAAGATTTGCTGCAAACTTC 60
|||||

QY 318 ATTAAGGAACATTCAGAGGAGCTCCAGGTGGTGGGAGGAGCAAGCTGTTTTAAG 377
|||||
Db 61 ATTAAGGAACATTCAGAGGAGCTCCATATGTGGTGGGAGGAGCAAGCTTTTAA 120
|||||

QY 378 TTATCTTCAGAAAGAGGTGCCATGCAATTTTCCCACTGATGTAAAGTCCCTCT 437
|||||
Db 121 TTATCTTCAGAAAGAGGTGCCATGCAATTTTCCCACTGATGTAAAGTCCCTCT 180
|||||

QY 438 GCTGCTGCGAGAGGAATTCCTTGAAGTGAATTAATCTGTGACACTTCAGACTG 497
|||||
Db 181 GCTGTGCGCCAGAGATTCCTTGAAGCTTAAATGATGCTGTGACACTTCAGACTG 240
|||||

QY 498 TACATTAATTAATTCCTCCAGAGGAGGTGCACTGCTCCAGAGCACTGTCCAGCAAT 557
|||||
Db 241 TATGTAATTAATTCCTCCAGAGGAGGTGCACTGCTCCAGCAATGCTGTTCAG----- 292
|||||

QY 558 CCAATTTGATCTATGACACCCCAACCAAGATATACGTCAACAGAGGAAATATGGA 617
|||||
Db 293 -----TATATGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 348
|||||

QY 618 ACTCCACCAAGAGATATGAGGCCCAACAGAGGAGATATGAGGCCCACTATGAGAT 677
|||||
Db 349 GCCCCACCTGAGAGATATGAGGCCCAACCTTAAAGAAATGAAAGCCGCCCTGTGAGATAC 408
|||||

QY 678 GGAAGCCCGCTGAGGAGATATGAGTCCCACTGGGGAGATATGAGTCCCACTGGGGAG 737
|||||
Db 409 AGAGCTCTCACTGTCAGATATGAGGCCCACTCTTGAATAGAGGCCCACTGTCAGAGGA 468
|||||

QY 738 TATGAGTCCCACTGAGGAGATATGAGGCCCACTGGGGAGATATGAGTCCCACTGGG 797
|||||
Db 469 TATGAGGCCCACTCTAGAGATATGAGGCCCACTCTTGAATATGAGGCCCACTCTCTC 528
|||||

QY 798 GGAATATGAGTCCCACTGGGGAGATATGAGGCCCACTGAGAGATATGAGGCCCACTCA 857
|||||
Db 529 GGAATATGAGGCCCACTCTGAGATATGAGGCCCACTGAGAGAAATGAGAGGCCCTCT 588
|||||

QY 858 GCTGGAATATGAGGCCCTCAACCCCTGCAATATGAGTCCATGCTGGAATATGAGCTGCC 917
|||||
Db 589 GCGGAAATACAGAGCTCTCACTGTCAGATCAAGAGCCAGGCTCTCAAGAAATCAAGAGCC 648
|||||

QY 918 TCTCAGAGATCTATGACAGCTCAGAGAGGAGCTCTCTCCACTACTCTCTCTTCTTCT 977
|||||
Db 649 -----CAGGCTCTGAAACGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 696
|||||

QY 978 GTCCATTACCACTCTCTCAGAGTAAACCTTGAAGACTCAACAGAGAG-GGACCTT 1036
|||||
Db 697 GTCCATTCTTAACCTCTTAAGATGAAACCTTGAAGACTCAACAGAGAGAGGTACCTT 756
|||||

QY 1037 AAAAAGTCAAGTCAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071
|||||
Db 757 AAAATGAGTCAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 791
|||||

RESULT 5
LOCUS BX296515 179222 bp DNA linear HTG 26-MAY-2003
DEFINITION Sus scrofa clone P198-121D21, *** SEQUENCING IN PROGRESS ***, 6
unordered pieces.
ACCESSION BX296515
VERSION BX296515.6 GI:31076160
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVIFIN; HTGS_DRAFT; HTGS_PULLTOP.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Cetartiodactyla; Suidae; Suidae; Sus.
REFERENCE 1 (bases 1 to 179222)

AUTHORS Tracey, A.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk; Clone requests: clonerequest@sanger.ac.uk
On May 26, 2003 this sequence version replaced GI:31043704.

COMMENT Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
Project name: BR121D21

Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 177502 bases at least Q40
Consensus quality: 177971 bases at least Q30
Consensus quality: 178283 bases at least Q20
Insert size: 178722; sum-of-contigs
Insert size: 167315; 12.1% error; agarose-fp
Quality coverage: 7.82x in Q20 bases; sum-of-contigs quality
coverage: 8.59x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 5850: contig of 5850 bp in length
* 5851
* 5950: gap of 100 bp
* 5951
* 74264: contig of 68314 bp in length
* 74265
* 74365: gap of 100 bp
* 74366
* 84656: contig of 10292 bp in length
* 84657
* 84756: gap of 100 bp
* 84757
* 108493: contig of 23737 bp in length
* 108494
* 108593: gap of 100 bp
* 108594
* 114454: contig of 5861 bp in length
* 114455
* 114554: gap of 100 bp
* 114555
* 179222: contig of 64668 bp in length.
Location/Qualifiers
1. 179222
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9823"
/clone="P198-121D21"
/clone_11b="P198"
1. 5850
/note="assembly fragment: 02838
fragment chain: 1"
5951. 74264
/note="assembly fragment: 01479
fragment chain: 1"
74365. 84656
/note="assembly fragment: 01113
fragment chain: 1"
84757. 108493
/note="assembly fragment: 00846
fragment chain: 1"
108594. 114454
/note="assembly fragment: 01058
fragment chain: 2"
114555. 179222
/note="assembly fragment: 02416
fragment chain: 2"
44880 c 44117 g 44831 t 502 others
BASE COUNT 44892 a 44880 c 44117 g 44831 t 502 others
ORIGIN
Query Match 24.6%; Score 347.2; DB 2; Length 179222;
Best Local Similarity 76.8%; Pred. No. 3.5e-75;

Matches 424; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 520 CTGCAGTGTCTCTCTCAAGACACCTTTCACAGATATCAATTTGATCTATGAGCCCC 579

DB 55353 CTCACAGATATCGAGAGCCCACTTCAGATATACGAGACCCCACTTCAGATATGAGAGCC 55412

QY 580 CACCACAGATATACAGTCCACCGAGGATATGAACTCCACAGAGATATGAG 639

DB 55413 CACTTCAGATATCGAGAGCCCACTTCAGAGATATGAGAGCCCACTTCAGATATGAG 55472

QY 640 CCCACAGAGGATATGAGAGCCCACTTCAGATATGAGAGCCCACTTCAGATATGAG 699

DB 55473 CCCATCTGCGGATATGAGAGCCCACTTCAGATATGAGAGCCCACTTCAGATATGAG 55532

QY 700 GAGTCCACCTGCGGATATGAGAGCCCACTTCAGATATGAGAGCCCACTTCAGATATGAG 759

DB 55533 GAGCTCCACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 55592

QY 760 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 819

DB 55593 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 55652

QY 820 GATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 879

DB 55653 GATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 55712

QY 880 CTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 939

DB 55713 TACATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 55772

QY 940 AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 999

DB 55773 AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 55832

QY 1000 GTTAAACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1059

DB 55833 AGTAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 55892

QY 1060 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071

DB 55893 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 55904

RESULT 6
BX470149 222469 bp DNA linear HTG 07-MAY-2003
LOCUS
DEFINITION
Dario rerio clone DKX-242K7, *** SEQUENCING IN PROGRESS ***, 50
unordered pieces.
BX470149
VERSION
BX470149.3 GI:30424228
KEYWORDS
HTG; HTGS PHASE1.
SOURCE
Dario rerio (zebrafish)
ORGANISM
Bkaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 222469)
Direct Submission
Submitted (06-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 7, 2003 this sequence version replaced gi:30387077.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zK242K7
----- Summary Statistics
Assembly program: XGAP4, version 4.5
Chemistry: Dye-terminator; 100% of reads

Consensus quality: 197958 bases at least Q40
Consensus quality: 206715 bases at least Q30
Consensus quality: 212351 bases at least Q20
Insert size: 217569; sum-of-contigs
Insert size: 165837; 6.2% error; agarose-fp
Quality coverage: 2.48x in Q20 bases; sum-of-contigs quality
coverage: 3.95x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
6657 6656: contig of 6656 bp in length
6657 6756: gap of 100 bp
6757 10255: contig of 3499 bp in length
10256 10355: gap of 100 bp
10356 13323: contig of 2868 bp in length
13324 13323: gap of 100 bp
13324 13935: contig of 6612 bp in length
13936 20035: gap of 100 bp
20036 22597: contig of 2562 bp in length
22598 22697: gap of 100 bp
22698 28569: contig of 5872 bp in length
28570 28670: gap of 100 bp
28670 37965: contig of 9236 bp in length
37966 38065: gap of 100 bp
38066 41140: contig of 3075 bp in length
41141 41240: gap of 100 bp
41241 45108: contig of 3868 bp in length
45109 45208: gap of 100 bp
45209 47730: contig of 2522 bp in length
47731 47831: gap of 100 bp
47832 51884: contig of 4054 bp in length
51885 51984: gap of 100 bp
51985 55957: contig of 3873 bp in length
55958 55957: gap of 100 bp
55958 60498: contig of 4541 bp in length
60499 60598: gap of 100 bp
60599 63641: contig of 3043 bp in length
63642 63741: gap of 100 bp
63742 72674: contig of 8933 bp in length
72675 72774: gap of 100 bp
72775 75602: contig of 2828 bp in length
75603 75702: gap of 100 bp
75703 78062: contig of 2360 bp in length
78063 78162: gap of 100 bp
78163 84485: contig of 6323 bp in length
84486 84585: gap of 100 bp
84586 87554: contig of 2968 bp in length
87555 92487: gap of 100 bp
92488 92587: contig of 4834 bp in length
92588 96104: gap of 100 bp
96105 96204: gap of 100 bp
96205 101479: contig of 5275 bp in length
101480 101579: gap of 100 bp
101580 104979: contig of 3400 bp in length
104980 105079: gap of 100 bp
105080 108003: contig of 2924 bp in length
108004 108103: gap of 100 bp
108104 113151: contig of 5046 bp in length
113152 113251: gap of 100 bp
113252 116585: contig of 3334 bp in length
116586 116685: gap of 100 bp
116687 119346: contig of 2661 bp in length
119347 119446: gap of 100 bp
119447 123125: contig of 3679 bp in length
123126 123225: gap of 100 bp
123226 131266: contig of 8041 bp in length

```
* 131267 131366: gap of 100 bp
* 131367 131693: contig of 4827 bp in length
* 136194 136293: gap of 100 bp
* 136294 139482: contig of 3189 bp in length
* 139483 139582: gap of 100 bp
* 139583 141875: contig of 2293 bp in length
* 141876 141975: gap of 100 bp
* 141976 144122: contig of 2147 bp in length
* 144123 144222: gap of 100 bp
* 146223 146368: contig of 2746 bp in length
* 146369 147068: gap of 100 bp
* 147069 149399: contig of 2331 bp in length
* 149400 149499: gap of 100 bp
* 149500 153963: contig of 4464 bp in length
* 153964 154063: gap of 100 bp
* 154064 156779: contig of 2516 bp in length
* 156780 156879: gap of 100 bp
* 156880 158928: contig of 2249 bp in length
* 158929 159028: gap of 100 bp
* 159029 164542: contig of 5514 bp in length
* 164543 164642: gap of 100 bp
* 164643 172054: contig of 7412 bp in length
* 172055 172154: gap of 100 bp
* 172155 175210: contig of 3056 bp in length
* 175211 175310: gap of 100 bp
* 175311 179379: contig of 4069 bp in length
* 179380 179479: gap of 100 bp
* 179480 182698: contig of 3219 bp in length
* 182699 182798: gap of 100 bp
* 182799 186779: contig of 3981 bp in length
* 186780 186879: gap of 100 bp
* 186880 195626: contig of 8747 bp in length
* 195627 195726: gap of 100 bp
* 195727 199396: contig of 3670 bp in length
* 199397 199496: gap of 100 bp
* 199497 206412: contig of 6916 bp in length
* 206413 206512: gap of 100 bp
* 206513 209619: contig of 3107 bp in length
* 209620 209719: gap of 100 bp
* 209720 219083: contig of 9364 bp in length
* 219084 219183: gap of 100 bp
* 219184 222469: contig of 3286 bp in length.
```

FEATURES

SOURCE

```
1..222469
/organism="Danio rerio"
/mol_type="Genomic DNA"
/db_xref="taxon:7955"
/clone="DKEX-242K7"
/clone_1fb="DanioKey"
1..6656
/note="assembly_fragment:01359
fragment_chain:1"
6757..10255
/note="assembly_fragment:01385
fragment_chain:1"
10356..13323
/note="assembly_fragment:00592
fragment_chain:1"
13324..19935
/note="assembly_fragment:00237
fragment_chain:1"
20036..22597
/note="assembly_fragment:00949
fragment_chain:1"
22598..28569
/note="assembly_fragment:00615
fragment_chain:1"
28570..37965
/note="assembly_fragment:01306
fragment_chain:2"
38066..41140
/note="assembly_fragment:00487
fragment_chain:2"
```

```
misc_feature 41241..45108
/note="assembly_fragment:00232
fragment_chain:2"
misc_feature 45209..47730
/note="assembly_fragment:00935
fragment_chain:2"
misc_feature 47831..51884
/note="assembly_fragment:01415
fragment_chain:2"
misc_feature 51985..55857
/note="assembly_fragment:00878
fragment_chain:3"
misc_feature 55958..60498
/note="assembly_fragment:01208
fragment_chain:3"
misc_feature 60599..63641
/note="assembly_fragment:01266
fragment_chain:3"
misc_feature 63742..72674
/note="assembly_fragment:01329
fragment_chain:3"
misc_feature 72775..75602
/note="assembly_fragment:01183
fragment_chain:4"
misc_feature 75703..78062
/note="assembly_fragment:01179
```

Query Match 24.6%; Score 347.2; DB 2; Length 222469;
Best Local Similarity 76.8%; Pred. No. 3.5e-75;
Matches 424; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```
QY 520 CTGAGAGTGTCTCTCAAGACACCTTTCAGATATCCAAATGTGATCTATGACCCC 579
DB 197009 CTCAGAGATTCGAGGCCCACTTCAGATATGAGAGGCCCACTTCAGATATGAGAGCCC 197068
QY 580 CACCACGAGATATATACAGTCCACACGAGGAGATATGAGATCCACAGAGATATGAGAG 639
DB 197069 CACTTCAGATATGAGAGGCCCACTTCAGATATGAGATCCACAGAGATATGAGAG 197128
QY 640 CCCAACGAGGAGATATGAGAGGCCCACTTCAGATATGAGATCCACAGAGATATGAGAG 699
DB 197129 CCCACTGCGGAGATATGAGAGGCCCACTTCAGATATGAGATCCACAGAGATATGAGAG 197188
QY 700 GAGTCCACCTGCGGAGATATGAGAGGCCCACTTCAGATATGAGATCCACAGAGATATGAGAG 759
DB 197189 GAGCTCACCTGAGAGACAGAGAGGCCCACTTCAGATATGAGATCCACAGAGATATGAGAG 197248
QY 760 ATGAGGCCCACTGCGGAGATATGAGAGGCCCACTTCAGATATGAGATCCACAGAGATATGAGAG 819
DB 197249 ATGAGGCCCACTGCGGAGATATGAGAGGCCCACTTCAGATATGAGATCCACAGAGATATGAGAG 197308
QY 820 GATATGAGGCCCACTGCGGAGATATGAGAGGCCCACTTCAGATATGAGATCCACAGAGATATGAGAG 879
DB 197309 GATATGAGGCCCACTGCGGAGATATGAGAGGCCCACTTCAGATATGAGATCCACAGAGATATGAGAG 197368
QY 880 CTGATATGAGAGCTTCATGCTGAGAAATACAGTGTCTTCACAGATCTATGAGACGCTC 939
DB 197369 TAGCATATGAGAGGCCCACTGCTGAGAAATAGAGTGTCTTCACAAATCTGTGACAGCCC 197428
QY 940 AGCAGAGATCTTCTTCCCACTACCTCATCTTCTTATAGTTCATTTACACCTTCTCAGA 999
DB 197429 AGCCGAGAGCTTCTTCTTCCCACTACCTCATCTTCTTATAGTTCATTTACACCTTCTCAGA 197488
QY 1000 GTTAACTTGAAGACTCATCAAGCAAGAGGAGCCCTTAAATCTGAAGTCAAGTATGAGAG 1059
DB 197489 AGTAACTTGAAGATTCACCAAGCAAGAGGATCTTAAATCTGAAGTCAAGTATGAGAG 197548
QY 1060 GAAGACCCAGGT 1071
DB 197549 GAGGACTCAGGT 197560
```

RESULT 7

HS250D10 220895 bp DNA linear PRI 05-JUN-2003
 LOCUS
 DEFINITION Human DNA sequence from clone CTA-250D10 on chromosome 22 contains the genes for SRBF2 (steroid regulatory element binding transcription factor 2), NGA (alpha-N-acetylglucosaminidase), a gene similar to neuronal-specific septin 3, a pseudogene similar to ANT2 (adenine nucleotide translocator 2), 2 mRNAs based on ESTs, a genomic marker D2S1178, a CA repeat polymorphism, ESTs and a Cpg island, complete sequence.
 Z99716
 VERSION Z99716.4 GI:4456457
 KEYWORDS HTG; ANT2; Cpg Island; D2S1178; NGA; septin 3; SRBF2.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 220895)
 AUTHORS Clark, G.
 JOURNAL Direct Submission
 Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On Mar 21, 1999 this sequence version replaced GI:4164339.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Bm: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 CTA-250D10 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218.
 VECTOR: pBAC1081
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk

 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
 This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22
 This sequence is the entire insert of clone CTA-250D10. The true left end of clone RP1-18601 is at 129979 in this sequence. The true left end of clone RP3-359016 is at 1335 in this sequence. The true right end of clone RP5-821D11 is at 23458 in this sequence. The true right end of clone RP3-359016 is at 118711 in this sequence.
 Location/Qualifiers
 1..220895
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="22"
 /clone="CTA-250D10"
 /clone_1b="CTG78SK-A1"
 17..29
 /note="2.6 copies 5 mer ATTGTG 26% conserved"
 repeat_region
 33..325
 /note="AluX repeat: matches 1..301 of consensus"
 repeat_region
 326..338

repeat_region
 667..683
 /note="2.1 copies 8 mer AGAAGACC 34% conserved"
 repeat_region
 740..751
 /note="2.4 copies 5 mer TTTTA 24% conserved"
 repeat_region
 1443..1457
 /note="7.5 copies 2 mer AC 21% conserved"
 repeat_region
 1715..1997
 /note="Alu repeat: matches 5..293 of consensus"
 repeat_region
 2103..2378
 /note="Alu repeat: matches 2..296 of consensus"
 repeat_region
 2844..2855
 /note="12.0 copies 1 mer A 24% conserved"
 repeat_region
 complement(2884..3033)
 /note="MIR repeat: matches 70..235 of consensus"
 repeat_region
 complement(3040..3344)
 /note="Alu repeat: matches 1..310 of consensus"
 repeat_region
 3446..3681
 /note="Alu repeat: matches 5256..5471 of consensus"
 repeat_region
 3746..3837
 /note="Alu repeat: matches 5615..5707 of consensus"
 repeat_region
 3838..4142
 /note="Alu repeat: matches 2..305 of consensus"
 repeat_region
 4143..4441
 /note="Alu repeat: matches 5707..5999 of consensus"
 repeat_region
 4442..4747
 /note="Alu repeat: matches 1..292 of consensus"
 repeat_region
 4748..4858
 /note="Alu repeat: matches 5999..6110 of consensus"
 repeat_region
 4881..5172
 /note="Alu repeat: matches 1..296 of consensus"
 repeat_region
 5418..5429
 /note="2.4 copies 5 mer GCTGA 24% conserved"
 repeat_region
 complement(5769..6021)
 /note="match: GSS: km:A0069839"
 repeat_region
 6074..6089
 /note="2.0 copies 8 mer CCCGCTG 32% conserved"
 repeat_region
 6139..6153
 /note="2.5 copies 6 mer TTTTA 21% conserved"
 repeat_region
 6141..6155
 /note="3.0 copies 5 mer TTTTA 21% conserved"
 repeat_region
 complement(6366..6572)
 /note="MIR repeat: matches 5..214 of consensus"
 repeat_region
 6839..6852
 /note="2.3 copies 6 mer GGAAT 28% conserved"
 repeat_region
 7088..7114
 /note="6.8 copies 4 mer TTTA 36% conserved"
 repeat_region
 7091..7109
 /note="1.9 copies 10 mer ATTATTTAC 38% conserved"
 repeat_region
 complement(7122..7418)
 /note="Alu repeat: matches 1..300 of consensus"
 repeat_region
 7464..7502
 /note="19.5 copies 2 mer AT 78% conserved"
 repeat_region
 7502..7524
 /note="11.5 copies 2 mer AC 46% conserved"
 repeat_region
 7523..7538
 /note="2.0 copies 8 mer CAAATAT 32% conserved"
 repeat_region
 7535..7563
 /note="14.5 copies 2 mer AT 49% conserved"
 repeat_region
 complement(7564..7859)
 /note="Alu repeat: matches 1..296 of consensus"
 repeat_region
 7873..7893
 /note="5.2 copies 4 mer TGT 33% conserved"
 repeat_region
 7914..8200
 /note="Alu repeat: matches 1..290 of consensus"
 repeat_region
 complement(8202..8318)
 /note="Alu/FLM repeat: matches 1..117 of consensus"
 repeat_region
 8387..8409
 /note="11.5 copies 2 mer AC 46% conserved"
 repeat_region
 8483..8498
 /note="2.0 copies 8 mer GTGTGT 32% conserved"
 repeat_region
 8486..8496
 /note="2.2 copies 5 mer GTGT 22% conserved"

```

repeat_region      8606..8622
                    /note="2.4 copies 7 mer AGAGAAA 34% conserved"
repeat_region      complement(8767..8899)
                    /note="MER115 repeat: matches 33..268 of consensus"
repeat_region      9004..9014
                    /note="2.2 copies 5 mer CTGAG 22% conserved"
repeat_region      9037..9140
                    /note="WIR repeat: matches 17..123 of consensus"
repeat_region      9509..9823
                    /note="AluXk repeat: matches 1..312 of consensus"
repeat_region      9872..9883
                    /note="2.4 copies 5 mer GTCTT 24% conserved"
gene               10071..99606
                    /gene="SRBP2"
mRNA               join(<10071..10520,11851..12032,14129..14275,17038..17259,
                    18577..18691,18783..18964,20469..20661,21182..21363,
                    23956..24232,28082..28251,36357..36522,38060..38177,
                    40292..40401,41889..42021,43570..43738,46210..46395,
                    48103..48214,48680..48900)
                    /gene="SRBP2"
                    /product="bK250D10.1 (sterol regulatory element binding
                    transcription factor 2)"
                    /note="sterol regulatory element binding transcription
                    factor 2"
match: cDNAs: Bm:U02031 Bm:U12330 Bm:U09068 Bm:U09103
Bm:U16995 Bm:U22818 Bm:U22819 Bm:U12329 Bm:AB017337
match: ESTs: Bm:R36312 Bm:AL040138 Bm:H92808 Bm:AA053886"
/evidence=not_experimental
join(<10071..10520,11851..12032,14129..14275,17038..17259,
18577..18691,18783..18964,20469..20661,21182..21363,
23956..24232,28082..28251,36357..36522,38060..38177,
40292..40401,41889..42021,43570..43738,46210..46395,
48103..48214,48680..48900)
/gene="SRBP2"
/note="match: proteins: Sw:Q12772 Sw:Q60429 Sw:P36956
Sw:Q60416 Tr:O15352"
/codon_start=3
/evidence=not_experimental
/product="bK250D10.1 (sterol regulatory element binding
transcription factor 2)"
/protein_id="CAB41234.1"
/db_xref="GI:5578715"
/translaction="MLOVSNVNGVRPRLPESRQCSSPRSGSGSSSGSSSSS
NGSSSSGAVDPVSQSTQVTLPSFSSAASPOAPTLQVVSPTSPTPRATPIID
PRPQPQPTOLQOQOTWITPTSTPTTIOQPLTYQAATSFOVLQVQSVLT
SSOVPTVIOOVQVQVQAVLTQVANGTLQTLAPVQVQVQVQVQVQVQVQVQV
KTDLVLTTLKDSPVAAVONPALTLITPTALQVQVQVQVQVQVQVQVQVQV
GQKVPKIQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
HKSGLAKLIDVITYIQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
DFQNVLMSPSPADSGSQAGSPSTSDSESPSLDDAKYKDEPDEFPVALGAVDSS
RILCVLFLCLSPNPLSLQWQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
LILMVGAVVLSVYVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
LQTCGAVGALPTSRDLQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
PDRKATSRAPALAVYRHQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
LVYIHLTAMGLKTRCGKGLFSLSPYLSRAQSLCGPSPHSAVPSILKRLCPIDQKPF

```

Query Match 20.0% Score 282.6; DB 9; Length 220895;
 Best Local Similarity 74.6% Pred. No. 3.7e-59;
 Matches 387; Conservative 0; Mismatches 119; Indels 13; Gaps 2;

```

QY 554 ATATCCATTGTGATCTATGAGCCGCCACACAGATATATACGTCACCAAGAGGAGATA 613
DB 169998 ATTCCTCAATTATGTCATGAGCCGCCACCTGACAGATATGAGCCCACTCCCGAGATA 170057
QY 614 TGAAGCTCCAGAGAGATATGAGCCCAACAGAGGAGATATGAGCCCACTTATGAG 673
DB 170058 CGAGCCCACTGAGAGATATGAGCCCAACCTTATGAGAGATATGAGCCCACTTATGAG 170117
QY 674 ATATGAGCCCGCTGTGAGATATGAGTCCACCTGAGGAGATATGAGATCCCACTTGG 733
DB 170118 ATACAGAGCTTCACTGTGAGATATGAGCCCACTTGTGAGATGAGAGCCCACTTGG 170177
QY 734 GGAATATGAGATCCCACTGAGGAGATATGAGCCCACTTGTGAGGAGATATGAGATCCCACT 793

```

```

DB 170178 AGAGTATGAGCCCACTTATGAGATATGAGAGCCCACTTGTGATATGAGAGCCCACT 170237
QY 794 TGGGGATATGAGTCCCACTTATGAGAGCCCACTTGTGAGATATGAGAGCCCACTTATGAG 853
DB 170238 TCTCGATATGAGAGCCCACTTATGAGATATGAGAGCCCACTTGTGAGATATGAGAGCC 170297
QY 854 ACAGAGTATGAGAGCCCACTTATGAGATATGAGAGCCCACTTGTGAGATATGAGAGCCCACT 913
DB 170298 GCTCGAGATATGAGAGCCCACTTATGAGATATGAGAGCCCACTTGTGAGATATGAGAG 170357
QY 914 TGCCTTACAGATATGAGAGCCCACTTATGAGAGCCCACTTGTGAGATATGAGAGCCCACT 973
DB 170358 AGCC-----CAGGCTCTTGAAGAGAGCCCTTCTTCCCTCTGCTCTTCTTCTTCTT 170405
QY 974 TTAGGCTATTTTACCACTTCTTGTGAGATATGAGAGCCCACTTGTGAGATATGAGAGCCCACT 1032
DB 170406 TCAAGCTATTTTACCACTTCTTGTGAGATATGAGAGCCCACTTGTGAGATATGAGAG 170465
QY 1033 CCTTAAATGAGATCACTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071
DB 170466 CCTTAAATGAGATCACTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 170504

```

RESULT 8

LOCUS BX296540/c 129624 bp DNA linear HTG 02-APR-2003
 DEFINITION unidirectional clone p18-231K18, *** SEQUENCING IN PROGRESS ***; 37
 unidirectional pieces.
 ACCESSION BX296540
 VERSION BX296540.2 GI:29500961
 KEYWORDS HTG; HTGS PHASE1.
 SOURCE Sus ecto1a (p19)
 ORGANISM Sus ecto1a

REFERENCE 1 (bases 1 to 129624)
 AUTHORS Burton, J.
 TITLE Direct Submission
 JOURNAL Submitted (02-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Apr 2, 2003 this sequence version replaced gi:129335441.

COMMENT

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: b7231K18
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 113740 bases at least Q40
 Consensus quality: 118187 bases at least Q30
 Consensus quality: 120712 bases at least Q20
 Insert size: 126024; sum-of-coverage
 Insert size: 154528; 4.6% error; agarose-ff
 Quality coverage: 2.35x in Q20 bases; sum-of-coverage
 coverage: 3.19x in Q20 bases; agarose-ff

----- NOTES: This is a 'working draft' sequence. It currently
 consists of 37 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 * 1 4262: contig of 4262 bp in length
 * 4263 4362: gap of 100 bp
 * 4363 6991: contig of 2629 bp in length
 * 6992 7091: gap of 100 bp

```
* 7092      9480: contig of 2389 bp in length
* 9481      9580: gap of 100 bp
* 9581      11688: contig of 2108 bp in length
* 11689     11788: gap of 100 bp
* 11789     14658: contig of 2870 bp in length
* 14659     14758: gap of 100 bp
* 14759     18440: contig of 3682 bp in length
* 18441     18540: gap of 100 bp
* 18541     20818: contig of 2278 bp in length
* 20819     20918: gap of 100 bp
* 20919     22969: contig of 2051 bp in length
* 22970     23069: gap of 100 bp
* 23070     27584: contig of 4515 bp in length
* 27585     27684: gap of 100 bp
* 27685     33211: contig of 5527 bp in length
* 33212     33311: gap of 100 bp
* 33312     35521: contig of 2210 bp in length
* 35522     38528: contig of 2907 bp in length
* 38529     38628: gap of 100 bp
* 38629     42063: contig of 3435 bp in length
* 42064     42163: gap of 100 bp
* 42164     44164: contig of 2001 bp in length
* 44165     44264: gap of 100 bp
* 44265     47707: contig of 3443 bp in length
* 47708     47807: gap of 100 bp
* 47808     50737: contig of 2930 bp in length
* 50738     50837: gap of 100 bp
* 50838     57066: contig of 6229 bp in length
* 57067     57166: gap of 100 bp
* 57167     59978: contig of 2812 bp in length
* 59979     60078: gap of 100 bp
* 60079     62560: contig of 2482 bp in length
* 62561     62660: gap of 100 bp
* 62661     64963: contig of 2303 bp in length
* 64964     65063: gap of 100 bp
* 65064     67191: contig of 2128 bp in length
* 67192     67291: gap of 100 bp
* 67292     72005: contig of 4714 bp in length
* 72006     72105: gap of 100 bp
* 72106     76595: contig of 4490 bp in length
* 76596     76695: gap of 100 bp
* 76696     79825: contig of 1130 bp in length
* 79826     79925: gap of 100 bp
* 79926     83540: contig of 3615 bp in length
* 83541     83640: gap of 100 bp
* 83641     86276: contig of 2636 bp in length
* 86277     86376: gap of 100 bp
* 86377     94620: contig of 8244 bp in length
* 94621     94720: gap of 100 bp
* 94721     97520: contig of 2800 bp in length
* 97521     97620: gap of 100 bp
* 97621     99784: contig of 2164 bp in length
* 99785     99884: gap of 100 bp
* 99885     102112: contig of 2228 bp in length
* 102113     102212: gap of 100 bp
* 102213     104242: contig of 2030 bp in length
* 104243     104342: gap of 100 bp
* 104343     110710: contig of 6368 bp in length
* 110711     110810: gap of 100 bp
* 110811     113367: contig of 2557 bp in length
* 113368     113467: gap of 100 bp
* 113468     115505: contig of 2038 bp in length
* 115506     115605: gap of 100 bp
* 115606     121874: contig of 6269 bp in length
* 121875     121974: gap of 100 bp
* 121975     127096: contig of 5122 bp in length
* 127097     127196: gap of 100 bp
* 127197     129624: contig of 2428 bp in length.
*
Location/Qualifiers
1..129624
/organism="Sus scrofa"
/mol_type="genomic DNA"
```

```
misc_feature      /db_xref="taxon:96823"
                  /clone="PigR-231X18"
                  /clone_1fb="PigR"
                  1..4262
                  /note="assembly_fragment:00361
                  fragment_chain:1"
                  4363..6591
                  /note="assembly_fragment:00108
                  fragment_chain:1"
                  7092..9480
                  /note="assembly_fragment:00460
                  fragment_chain:1"
                  9581..11688
                  /note="assembly_fragment:00959
                  fragment_chain:1"
                  11789..14658
                  /note="assembly_fragment:00995
                  fragment_chain:2"
                  14759..18440
                  /note="assembly_fragment:00608
                  fragment_chain:2"
                  18541..20818
                  /note="assembly_fragment:01101
                  fragment_chain:2"
                  20919..22969
                  /note="assembly_fragment:00996
                  fragment_chain:2"
                  23070..27584
                  /note="assembly_fragment:00217
                  fragment_chain:3"
                  27685..33211
                  /note="assembly_fragment:00019
                  fragment_chain:3"
                  33312..35521
                  /note="assembly_fragment:00548
                  fragment_chain:3"
                  35622..38528
                  /note="assembly_fragment:00815
                  fragment_chain:4"
                  38629..42063
                  /note="assembly_fragment:00246
                  fragment_chain:4"
                  42164..44164
                  /note="assembly_fragment:00859
                  fragment_chain:5"
                  44265..47707
                  /note="assembly_fragment:00964
                  fragment_chain:5"
                  47808..50737
                  /note="assembly_fragment:01192
                  fragment_chain:6"
                  50838..57066
                  /note="assembly_fragment:00542
                  fragment_chain:6"
                  57167..59978
                  /note="assembly_fragment:01304
                  fragment_chain:7"
                  60079..62560
                  /note="assembly_fragment:00948
                  fragment_chain:7"
                  62661..64963
                  /note="assembly_fragment:00016"
                  65064..67191
                  /note="assembly_fragment:00065"
                  67292..72005
                  /note="assembly_fragment:00070"
                  72106..76595
                  /note="assembly_fragment:00136"
                  76696..79825
                  /note="assembly_fragment:00162"
                  79926..83540
                  /note="assembly_fragment:00367"
                  83641..86276
```

```

misc_feature      /note="assembly_fragment:00492"
6377..94620
/note="assembly_fragment:00545"
misc_feature      /note="assembly_fragment:00545"
94721..97520
/note="assembly_fragment:00745"
misc_feature      /note="assembly_fragment:00801"
97621..99784

Query Match      15.8%; Score 223.2; DB 2; Length 129624;
Best local similarity 74.2%; Pred. No. 2e-44;
Matches 282; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 550 CAGCATATCCATGTGATCTATATGAGCCGCCACGAGATATACAGTCCAAACGAGGG 609
DB 65511 CATCATGCCCAAGTGTGTATGATGACCCCTACGATGAGATGAGAGCCCAACGAG 65452
QY 610 AATATGAACTTCACAGAGATATGAGCCCAACGAGGGGATATGAGAGCCCACTTA 669
DB 65451 AATATGAGCCCACTTCAGAGATATGAGAGCCCACTTCAGAGATATGAGAGCCCACTTC 65392
QY 670 TGGATATGAGAGCCCGCTGTGGATATGAGTCCCACTGGGGATATGAGTCCAC 729
DB 65391 CAGGATAGGAGCCCACTTCAGAGATATGAGAGCCCACTTCAGAGATATGAGAGCCCACT 65332
QY 730 CTGGGGGATATGAGATCCCACTGGGGGATATGAGAGCCCACTGGGGGATATGAGATCC 789
DB 65331 CTGAAAGACTGAAACCCCAACGAGATATGAGAGCCCACTGGGGATATGAGAGCTT 65272
QY 790 CACCTGGGGGATATGATGATCCCACTGGGGGATATGAGAGCCCACTTCAGAGATATGAGAG 849
DB 65271 CACCTGGGGGATATGATGATCCCACTGGGGGATATGAGAGCCCACTTCAGAGATATGAGAG 65212
QY 850 CCCCACGAGTGAATGAAAGCCCTCACTGATATGAGATCTTCATCTCTGGAATA 909
DB 65211 CCCCACGAGGATATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 65152
QY 910 CAGCTGCTCTTCACGATCT 929
DB 65151 ATGCCCGCTTCAGAGATAT 65132

RESULT 9
AC113593      168425 bp      DNA      linear      HTG 06-JUN-2002
LOCUS      Mus musculus clone RP23-363124, WORKING DRAFT SEQUENCE, 18 ordered
DEFINITION      pieces.
AC113593      GI:21327432
VERSION      HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
1 (bases 1 to 168425)
REFERENCE      Birren, B., Linton, L., Nusbaum, C. and Lander, E.
AUTHORS      Mus musculus, clone RP23-363124
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 168425)
AUTHORS      Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barna, N., Baetien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagoe, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kelle, C., Larocque, K., Lamazares, R.,
Lander, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McKernan, K., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunhahang, P., Pierre, N., Pollara, V., Raymond, C.,

```

```

TITLE      JOURNAL
REFERENCE      Submitted (03-MAR-2002) Whitehead Institute/MIT Center for Genome
AUTHORS      Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 168425)
Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barna, N., Baetien, V., Bloom, T., Boguslavsky, L.,
Boukhalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,
Chazaro, B., Chapel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagoe, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kelle, C., Larocque, K., Lamazares, R.,
Lander, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McKernan, K., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunhahang, P., Pierre, N., Pollara, V., Raymond, C.,
Young, G., Zainoun, D., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gi:21313869.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WTHR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23816
Center clone name: 363.1.24
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 162227 bases at least Q40
Consensus quality: 164865 bases at least Q30
Consensus quality: 165972 bases at least Q20
Insert size: 16200; agarose-1p
Insert size: 166725; sum-of-ctnigs
Quality coverage: 8.6 in Q20 bases; agarose-1p
Quality coverage: 8.3 in Q20 bases; sum-of-ctnigs
-----
NOTES: This is a 'working draft' sequence. It currently
* consists of 18 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 1412: contig of 1412 bp in length
1413 1512: gap of 100 bp
1513 3366: contig of 1854 bp in length
3367 3466: gap of 100 bp

```


	Mouse	BAC
misc_feature	1. .1412	/clone_11b="RPci-23 Female
misc_feature	/note="assembly_fragment"	
misc_feature	1513. .3366	
misc_feature	/note="assembly_fragment"	
misc_feature	3467. .5542	
misc_feature	/note="assembly_fragment"	
misc_feature	5643. .7949	
misc_feature	/note="assembly_fragment"	
misc_feature	8050. .42610	
misc_feature	/note="assembly_fragment"	
misc_feature	42711. .45717	
misc_feature	/note="assembly_fragment"	
misc_feature	45818. .50106	
misc_feature	/note="assembly_fragment"	
misc_feature	50207. .52945	
misc_feature	/note="assembly_fragment"	
misc_feature	53046. .56269	
misc_feature	/note="assembly_fragment"	
misc_feature	56370. .59897	
misc_feature	/note="assembly_fragment"	
misc_feature	59998. .67207	
misc_feature	/note="assembly_fragment"	
misc_feature	67308. .78743	
misc_feature	/note="assembly_fragment"	
misc_feature	78844. .89991	
misc_feature	/note="assembly_fragment"	
misc_feature	90092. .107918	
misc_feature	/note="assembly_fragment"	
misc_feature	108019. .124874	
misc_feature	/note="assembly_fragment"	
misc_feature	124975. .142093	
misc_feature	/note="assembly_fragment"	
misc_feature	142194. .167111	
misc_feature	/note="assembly_fragment"	
misc_feature	167412. .168425	

REFERENCE	4 (bases 1 to 224086)
AUTHORS	Jiang, H., Song, L. and Roe, B.A.

TITLE Direct Submission
JOURNAL Submitted (14-MAY-2003) Department of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
REFERENCE 5 (bases 1 to 224086)
AUTHORS Jiang,H., Song,L. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (30-MAY-2003) Department of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
COMMENT On May 30, 2003 this sequence version replaced gi:29124187.
 ----- Genome Center
 The University Of Oklahoma
 The University Of Oklahoma
 Center code:UOKNOR

FEATURES
 source
 1. 224086
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /map="15"
 /clone="T23-204m3"
 /clone_1b="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"
BASE COUNT 55592 a 55124 c 56095 g 57275 t
ORIGIN
 Query Match 13.4%; Score 188.8; DB 10; Length 224086;
 Best Local Similarity 65.2%; Pred. No. 6.8e-36;
 Matches 313; Conservative 0; Mismatches 157; Indels 10; Gaps 2;
 Oy 602 ACCAGGGGATATGGAATCTCCACAGAGATATGAGCCCAAGGGGGATATGAGC 661
 Db 162848 AGCTGTGAGATATGCTCCCTCCCTCTCTATATATGATGATACCCCAATGGGCTATGAGAT 162907
 Oy 662 CCCACTATGATATGAGAGCCCGCTGTGATATATGAGATCCCACTGGGGATATGAG 721
 Db 162908 TCCACCTCTGATATGAGCTCCACCTATGATATGAGATCCCACTCTCTGATATGAG 162967
 Oy 722 AGTCCCACTGGGGATATGAGATCTCCACCTGGGGATATGAGAGCCCACTGGGGATAT 781
 Db 162968 AGCCCACTATGAGATATGAGAGCCCACTCCAGATATGAGAACTCACTATGAGATC 163027
 Oy 782 TGAAGTCCCACTGGGGATATGAGATCTCCACCTGGGGATATGAGAGCCCACTGGAGG 841
 Db 163028 CGGTTCCTCCCACTCCAGATATGAGAGCCCGGCTATGAGGATCTCTCTCAAGG 163087
 Oy 842 ATATGAGAGCCCACTCCAGATATGAGAGCCCACTCCAGATATGAGAGCCCACTGGC 901
 Db 163088 GCGTATGATCACTCTCTGATATGAGAGCCCACTCTGATATGAGAGAGCTCACTGC 163147
 Oy 902 TGAATAATAGAGTCTCTCAAGATCTATGAGAGCTCAAGCACTTCAAGCTTC 952
 Db 163148 TGGGTCTGAAGAGGCTATCTCATGTCTGTGAGAGCTCAAGATCTCAAGCTTC 163207
 Oy 953 TCTTCCCACTCACTATCTTCTTGAAGTCAATTACCACTTCTCAAGATTAACCTTGA 1012
 Db 163208 TTTTCCCTCTACCTGCTCTCAAGATCACTCTCCGCTCTGAAGATTAACCTTGA 163267
 Oy 1013 GACTCAGCAAGC-AAAGGGCACTTAAATGAAGTCAAGTCAAGTAAAGAAAGCAAGT 1071
 Db 163268 GTTTCACCAAGCAAGCAAGCACTATATGTGAAGTCAAGTAAAGCAAGCACTTAAGT 163327

RESULT 11
 AC107527 253149 bp DNA linear HTG 13-MAY-2003
 LOCUS AC107527
 DEFINITION Rattus norvegicus clone CH230-92M24, WORKING DRAFT SEQUENCE, 3
 unnumbered pieces.
 AC107527
 AC107527.5 GI:30580771
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE
ORGANISM Rattus norvegicus (Norway rat)
REFERENCE 1 (bases 1 to 253149)
AUTHORS Munzy,D,Marie., Metker,M,lee., Abramson,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Alebrooks,S., Amin,A., Angiano,D.,
 Anyalebechi,V., Ayvagi,A., Ayodeji,M., Bacc,B., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barneshead,M., Benahmed,F.,
 Bernalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,B.,
 Cardenas,V., Carter,K., Cavazos,I., Cesarin,H., Chen,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Dugan-Rocha,S., Dunn,A., Dublin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Faller,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Genta,R., Garcia,A., Garner,T., Garza,M.,
 Georgegeorgis,B., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,M.,
 Gunaratne,P., Haaland,W., Hamli,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hilyk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Jorpechthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kovals,C., Kraft,C.T., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lounsbury,L., Louisedge,H., Lozada,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindartne,M., Mahmood,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapue,P., Martin,K., Martin,R., Martinez,B.,
 Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,
 Milosavljevic,A., Miner,G., Minj,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Muidasa,M., Murphy,M., Nair,L.,
 Narkervia,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
 Nwokilemech,O., Ohnomu,G., Olanmuogson,A., Pal,S., Parke,K.,
 Pasternak,S., Paul,H., Perez,A., Perez,J., Plankoch,C.,
 Plopper,F., Polndexter,A., Popovic,D., Primus,B., Pu,L., L.,
 Puzos,M., Qutroz,J., Rachlin,B., Reeves,K., Regier,M.A., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
 Shetty,J., Shvartsbeyn,A., Slason,I., Sitter,C.D., Smajls,D.,
 Sneed,A., Sodergren,B., Song,X.-Z., Sorelle,R., Sosa,J.,
 Steimle,M., Strong,R., Sutton,A., Swatke,A., Taboe,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejo,Z., Uemami,K.,
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willson,R., Wleczyk,R., Wodden,H., Worley,V.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinrock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 253149)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 253149)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
COMMENT The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described

In the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GKFL

Center clone name: CH230-92M24

Assembly program: Atlas 3.0:

Consensus quality: 242817 bases at least Q40

Consensus quality: 245517 bases at least Q30

Consensus quality: 247516 bases at least Q20

Estimated insert size: 257092; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

----- Summary Statistics

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 250781: contig of 250781 bp in length
 * 250782 250881: gap of unknown length
 * 250882 251966: contig of 1085 bp in length
 * 251967 252066: gap of unknown length
 * 252067 253149: contig of 1083 bp in length.

FEATURES

source

1. 253149

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-92M24"

2794..3331

/note="clone_boundary"

clone_end:T7

site:ECORI

end_sequence: BH296484"

2825..3625

/note="clone_boundary"

clone_end:T7

site:ECORI

end_sequence: BH296484"

complement(246782..247500)

/note="clone_boundary"

clone_end:Spf

site:ECORI

end_sequence: BH296486"

61494 C 61630 G 62235 T 3750 others

BASE COUNT 64040 a 61494 c 61630 g 62235 t 3750 others

ORIGIN

Query Match

Best Local Similarity 62.9%; Pred. No. 7.1e-32; Length 253149;

Matches 304; Conservative 0; Mismatches 169; Indels 10; Gaps 2;

Qy 602 ACCAGGGAATATGAACTCCACCAAGATATGAGCCCAACCAAGGAGATATGAGC 661
 Db 117968 AGCTGCGAGATATGATCCCTCTCTCTATAGTATGATACCCCTATGAGGCTATGAGT 117909

Qy 662 CCCACTATATGAGATATGAGCCCGCTGTGAGATATGAGTCCACTGTGGGAGATATG 721
 Db 117908 CCCACTCTGTGATATGAGCCCGCACTGTGATATGAGTCCACTGTGGGAGATATG 117849
 Qy 722 AGTCCACTGTGGGAGATATGAGTCCCGCTGTGAGATATGAGTCCCGCTGTGGGAGATATG 781
 Db 117848 AGTCCACTGTGGGAGATATGAGTCCCGCTGTGAGATATGAGTCCCGCTGTGGGAGATATG 117789
 Qy 782 TGAAGTCCACTGTGGGAGATATGAGTCCCGCTGTGAGATATGAGTCCCGCTGTGGGAGATATG 841
 Db 117788 CGATTCCTCCACTGTGGGAGATATGAGTCCCGCTGTGAGATATGAGTCCCGCTGTGGGAGATATG 117729
 Qy 842 ATATGAGGAGCCCAAGCAGTGTGAGATATGAGTCCCGCTGTGAGATATGAGTCCCGCTGTGGGAGATATG 901
 Db 117728 GAGTATGATATATATCTCTGTGATATGAGTCCCGCTGTGAGATATGAGTCCCGCTGTGGGAGATATG 117669
 Qy 902 TGAAGTATATGAGTGTCTCTGTGAGATATGAGTCCCGCTGTGAGATATGAGTCCCGCTGTGGGAGATATG 952
 Db 117668 TGAAGTATATGAGTGTCTCTGTGAGATATGAGTCCCGCTGTGAGATATGAGTCCCGCTGTGGGAGATATG 117609
 Qy 953 TCTTCCACTATCTATCTTTTGAAGTATATGAGTCCCGCTGTGAGATATGAGTCCCGCTGTGGGAGATATG 1012
 Db 117608 TTTCCCTCTACTCTGTGTGAGATATGAGTCCCGCTGTGAGATATGAGTCCCGCTGTGGGAGATATG 117549
 Qy 1013 GACTCAGCAGC-AAAAGGCACTTAAATGAGTATGAGTCCCGCTGTGAGATATGAGTCCCGCTGTGGGAGATATG 1071
 Db 117548 GTTTCAGCAGC-AAAAGGCACTTATGAGTATGAGTCCCGCTGTGAGATATGAGTCCCGCTGTGGGAGATATG 117489
 Qy 1072 GCC 1074
 Db 117488 GTC 117486

RESULT 12
 AC132969 270171 bp DNA linear HTG 20-NOV-2002
 LOCUS Rattus norvegicus clone CH230-327L20, WORKING DRAFT SEQUENCE.
 DEFINITION AC132969
 ACCESSION AC132969
 VERSION AC132969.3 GI:25319203
 KEYWORDS HTG; HTGS; PHAS2; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 270171)
 Muzny,D,Marie,,Metzker,M,Lea,,Abramson,S,,Adams,C,,Alder,J,,
 Allen,C,,Allen,H,,Altschuler,S,,Amin,A,,Anguiano,D,,
 Anyalebech,V,,Aoyagi,A,,Ayodeji,M,,Baca,R,,Baden,H,,
 Baldwin,D,,Bandaranaike,D,,Barber,M,,Barnstead,M,,Benahmed,F,,
 Biswal,K,,Blair,J,,Blankenburg,K,,Blyth,P,,Brown,M,,
 Bryant,N,,Bunay,C,,Burch,P,,Burrell,K,,Calderson,E,,
 Cardenas,V,,Carter,K,,Cavazos,I,,Cesari,H,,Center,A,,
 Chacko,J,,Chavez,D,,Chen,G,,Chen,R,,Chen,Y,,Chen,Z,,Chu,J,,
 Cleveland,C,,Cockrell,R,,Cox,C,,Coyler,M,,Crease,A,,D'Souza,L,,
 Davila,M,L,,Davis,C,,Davy-Carroll,L,,De Anda,C,,Dederich,D,,
 Delgado,O,,Denson,S,,Deramo,C,,Ding,Y,,Dinh,H,,Divya,K,,
 Draper,H,,Dugan-Rocha,S,,Dunn,A,,Durbin,K,,Duvall,B,,Eaves,K,,
 Egan,A,,Escotto,M,,Eugene,C,,Evans,C,A,,Falls,T,,Fan,G,,
 Fernandez,S,,Finley,M,,Flagg,N,,Forbes,L,,Foster,M,,Foster,P,,
 Frazer,C,M,,Gabriel,A,,Ganta,R,,Garcia,A,,Garner,T,,Garza,M,,
 Gebregeorgis,B,,Geer,K,,Gill,R,,Girdy,A,,Guerre,M,,Guevara,W,,
 Gunaratne,P,,Haaland,W,,Hamil,C,,Hamilton,C,,Hamilton,K,,
 Harvey,Y,,Havlik,P,,Hawes,A,,Henderson,N,,Hernandez,J,,
 Hernandez,R,,Hines,S,,Hladun,S,L,,Hodgson,A,,Hogues,M,,
 Hollins,B,,Howells,S,,Hulik,S,,Hume,J,,Idlebird,D,,Jackson,A,,
 Jackson,L,,Jacob,L,,Jiang,H,,Johnson,B,,Johnson,R,,Jolyet,A,,
 Karpathy,S,,Kelly,S,,Kelly,S,,Khan,Z,,King,L,,Kovar,C,,
 Kowals,C,,Kraft,C,L,,Lebow,H,,Levan,J,,Lewis,L,,Li,Z,,Liu,J,,
 Liu,J,,Liu,W,,Liu,Y,,London,P,,Longacre,S,,Lopez,J,,
 Lorensen,L,,Louiegeed,H,,Lozano,R,,Lu,X,,Ma,J,,
 Maheshwari,M,,Mahindaratne,M,,Mahmoud,M,,Malloy,K,,Mangum,A,,
 Mangum,B,,Mapua,P,,Martin,K,,Martin,R,,Martinez,B,,

Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundaas, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Pawlowski, O., Okunolu, G., Olarinmoye, A., Pal, S., Park, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Polidexter, A., Popovic, D., Potts, R., Pu, L., Pu, M., Quirroz, J., Rachlin, B., Reeves, K., Reider, M.A., Reigh, R., Rivas, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, B., Song, X.-Z., Sorrelle, R., Soza, J., Steidle, M., Strong, R., Sutcliffe, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejo, Z., Umant, K., Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Williams, R., Wleczek, R., Wood, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
Rat Genome Sequencing Consortium.
Unpublished

REFERENCE
2 (bases 1 to 270171)
Rat Genome Sequencing Consortium.
Submitted (05-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

AUTHORS
3 (bases 1 to 270171)
Rat Genome Sequencing Consortium.
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT
On Nov 20, 2002 this sequence version replaced gi:23908490.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----
Center project name: K80D
Center clone name: CH230-327L20

----- Summary Statistics -----
Assembly program: Phrap, version 0.990329
Consensus quality: 251525 bases at least Q40
Consensus quality: 254049 bases at least Q30
Consensus quality: 255599 bases at least Q20
Estimated insert size: 259926/ sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have

* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 270171: contig of 270171 bp in length.
Location/Qualifiers
1..270171
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-327L20"
1..2686
/note="igs end extension
clone end:5p6"
4191..5073
/note="clone boundary
clone end:5p6
site:
end sequence:B2160678"
complement(267941..268827)
/note="clone boundary
clone end:t7
site:
end sequence:B2160655"
B2160655
67733 a 63574 c 62822 g 63249 t 12793 others
ORIGIN

Query Match 12.2% Score 172.6; DB 2; Length 270171;
Best Local Similarity 62.9%; Pred. No. 7.1e-32;
Matches 304; Conservative 0; Mismatches 169; Indels 10; Gaps 2;

QY 602 ACCAGGGGAATATGGAAGTCCACGAGAAAGATATGAGCCCAACGAGGGGATATGAGC 661
DB 15216 AGTCGCGAATATGATATCCCTCTCTCTATATGATATGATATGATATGAGT 15157
QY 662 CCAACCTATGGAATATGAGCCCGCCCTGATATGAGATATGAGATATGAGT 721
DB 15156 CCAACCTCTGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGT 15097
QY 722 AGTCCCACTGAGGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGT 781
DB 15096 AGCCCACTATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGT 15037
QY 782 TGAAGTCCCACTGAGGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGT 841
DB 15036 CGGTCCCACTCCCGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGT 14977
QY 842 ATATGAGAGCCCACTGAGTATGAGATATGAGATATGAGATATGAGATATGAGATATGAGT 901
DB 14976 GATGATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAGT 14917
QY 902 TGAATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGT 952
DB 14916 TGGATGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGT 14857
QY 953 TCTTCCCACTATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAGT 1012
DB 14856 TTTCCCTCTATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAGT 14797
QY 1013 GATCAGCAAGC-AAAAGGCAAGCTTAAAGTATGAGATATGAGATATGAGATATGAGATATGAGT 1071
DB 14796 GTTTCACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGT 14737
QY 1072 GCC 1074
DB 14736 GTC 14734

RSRUT 13
AF499026
LOCUS AF499026 793 bp mRNA linear ROD 23-APR-2002
DEFINITION Rattus norvegicus Wt-domain binding protein 2 (Wbp2) mRNA, complete
CDB.

ACCESSION AF499026
 VERSION AF499026.1 GI:20269980
 KEYWORDS
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 793)
 AUTHORS Zannini, M. and Nitsch, R.
 TITLE Isolation of the cDNA encoding for rat Wbp2
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 793)
 AUTHORS Zannini, M. and Nitsch, R.
 TITLE Direct Submission
 JOURNAL Submitted (04-APR-2002) DBPCN, University of Naples Federico II, Via Pansini 5, Naples, NA 80131, Italy
 FEATURES
 source
 1..793
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 gene
 1..793
 /gene="Wbp2"
 CDS
 1..789
 /gene="Wbp2"
 /codon_start=1
 /product="WW-domain binding protein 2"
 /protein_id="AA18132.1"
 /db_xref="GI:20269981"
 /translation="MALNKHSGGGVIVNNTESILMSYDHYELFENDKQVPEAFK
 TKKGYVLTPEYVITPLSKGDAMQSPMPTLMDCRQKQVPGANPTKGYKABG
 GMBGSASTLTFTAGATIFGQRLQVASQSRGVENKATGYPTMGAVPPPPVA
 NGMYPCPGYPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
 AAKAKAAABAAAYVNPNGNPHVMTSQPPPPPPPPPPDKTQ"
 BASE COUNT 189 a 259 c 196 g 149 t
 ORIGIN
 Query Match 9.1%; Score 128.4; DB 10; Length 793;
 Best Local Similarity 60.3%; Pred. No. 7.5e-21;
 Matches 251; Conservative 0; Mismatches 156; Indels 9; Gaps 2;
 Db 1 ATGGCGCTCAACAGATCACTCAAGAGCGCGGAGTATCTCAACACACT--GAG 57
 Qy 36 ATGGCAGTGAACAGAGCCACCGAGAGCGTCTGAGGAGCCCTCATCCCTCTGAGCA 95
 Db 1 ATGGCGCTCAACAGATCACTCAAGAGCGCGGAGTATCTCAACACACT--GAG 57
 Qy 96 AGTGTCTTGAAGCAGTGTGAGATGTGACCTCTGCTTCTTCAAGAAACCAAGTGAATCC 155
 Db 58 AGCATCTTAATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 117
 Qy 156 TATCTCTTAAATGACCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 215
 Db 118 GAGGCGCTTCAAGAGGACCAAGAAAGGACCGTCTACCTTAACCTCATACCGGGCTATCTTT 177
 Qy 216 GTGACTTCACTTAATGATGAAGCCAGTCTTCTTCTTAAATGAGCGTTTGCGCTGAG 275
 Db 178 CTGTCT-----CAAGGCAAGATGCGATCGGCTCTCAATGATGCGCTTCACTGAG 231
 Qy 276 AGTGTCTTGAAGCAGTGTGAGATGTGACCTCTGCTTCTTCAAGAAACCAAGTGAATCC 335
 Db 232 AAGGACTGTGAGTCAAGAGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 291
 Qy 336 GCACTTCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 395
 Db 292 GCTGAAGCAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 351
 Qy 396 GGTGCAATGAAATTTGCCCACTGATGATGATGATGATGATGATGATGATGATGATGAT 451
 Db 352 GGTGCAATGAAATTTGCCCACTGATGATGATGATGATGATGATGATGATGATGATGAT 407

LOCUS MMU40826 1802 bp mRNA linear ROD 04-MAR-2003
 DEFINITION Mus musculus WW-domain binding protein 2 mRNA, complete cds.
 ACCESSION U40826
 VERSION U40826.1 GI:1777578
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1802)
 AUTHORS Sudol, M., Chen, H.I., Bougeret, C., Binford, A. and Bork, P.
 TITLE Characterization of a novel protein-binding module--the WW domain
 JOURNAL FASEB Lett. 369 (1), 67-71 (1995)
 MEDLINE 95369475
 PubMed 7641887
 REFERENCE 2 (bases 1 to 1802)
 AUTHORS Chen, H.I. and Sudol, M.
 TITLE The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 92 (17), 7813-7823 (1995)
 MEDLINE 95372370
 PubMed 7644498
 REFERENCE 3 (bases 1 to 1802)
 AUTHORS Chen, H.I. and Sudol, M.
 TITLE Identification and Characterization of Protein Ligands to the WW Domain by Western Ligand Blotting
 JOURNAL (in) Marshak, D.R. (Ed.), TECHNIQUES IN PROTEIN CHEMISTRY VII: 3-12; Academic Press, Inc., USA (1996)
 REFERENCE 4 (bases 1 to 1802)
 AUTHORS Chen, H.I. and Sudol, M.
 TITLE Direct Submission
 JOURNAL Submitted (17-NOV-1995) Henry I. Chen, Laboratory of Molecular Oncology, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA
 FEATURES
 source
 1..1802
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /dev_stage="16 day embryo"
 CDS
 40..825
 /note="WBP-2; putative protein ligand to the WW domain of Yes-kinase associated protein YAP"
 /codon_start=1
 /product="WW-domain binding protein 2"
 /protein_id="BAB40893.1"
 /db_xref="GI:1777579"
 /translation="MALNKHSGGGVIVNNTESILMSYDHYELFENDKQVPEAFK
 TKKGYVLTPEYVITPLSKGDAMQSPMPTLMDCRQKQVPGANPTKGYKABG
 GMBGSASTLTFTAGATIFGQRLQVASQSRGVENKATGYPTMGAVPPPPVA
 NGMYPCPGYPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
 AAKAKAAABAAAYVNPNGNPHVMTSQPPPPPPPPPPDKTQ"
 BASE COUNT 396 a 565 c 437 g 404 t
 ORIGIN
 Query Match 9.1%; Score 128.4; DB 10; Length 1802;
 Best Local Similarity 60.3%; Pred. No. 7.4e-21;
 Matches 251; Conservative 0; Mismatches 156; Indels 9; Gaps 2;
 Db 36 ATGGCAGTGAACAGAGCCACCGAGAGCGTCTGAGGAGCCCTCATCCCTCTGAGCA 95
 Qy 40 ATGGCGCTCAACAGATCACTCAAGAGCGCGGAGTATCTCAACACACT--GAG 96
 Db 96 AGTGTCTTGAAGCAGTGTGAGATGTGACCTCTGCTTCTTCAAGAAACCAAGTGAATCC 155
 Qy 97 AGCATCTTAATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 156
 Db 156 TATCTCTTAAATGACCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 215
 Qy 157 GAGGCGCTTCAAGAGGACCAAGAAAGGACCGTCTACCTTAACCTCATACCGGGCTATCTTT 216

Qy 216 GTGACTTCACTTAGTCATGATGACCCATGCTTCTTTATGATGCCGTTGGCCGTGATG 275
Db 217 CTGTCTAAG-----GGAGAGGAGCCCATGACAGTCTTCTCATGATGCCCTTTACTGATG 270
Qy 276 AGTACTGACACCTGGAACACCAATTTTGGCCCACTACATTAAAGAACACTGAG 335
Db 271 AAGGACTGTGAGATTAGACAGCCGCTGTTGTGTGAACTTATTAGGATAGTGAAG 330
Qy 336 GCAAGTCAGGATGATGCTGGGAGAGACAGCTGTTTATGTTATCTTCCAGAAAGGA 395
Db 331 GCTGAGAGAGAGGAGGCTGGAGAGGCTCCCTCTTCAAGCTACCTTCAAGAGAGG 390
Qy 396 GATGCGCATGCAATTTGCCCACTGATGTAAGTAAGCTGCTGCTGACAGAG 451
Db 391 GCGCGCATGATGTTGGGAGAGAGATCTCCAGAGTGATCTCAAGCTCCAGAGG 446

RESULT 15
BX470149/c
LOCUS
DEFINITION BX470149 222469 bp DNA linear HTG 07-MAY-2003
Dario rerio clone DKEX-242K7, *** SEQUENCING IN PROGRESS ***, 50
underead pieces.
ACCESSION BX470149
VERSION BX470149.3 GI:30424228
KEYWORDS HTG; HTGS PHASE1.
SOURCE Dario rerio (zebrafish)
ORGANISM Dario rerio
Bukarjota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Osteichthyes;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 222469)
DIRECT SUBMISSION
Submitted (05-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 7, 2003 this sequence version replaced gi:30387077.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: ZK242K7
----- Summary Statistics
Assembly program: XGAP; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 197958 bases at least Q40
Consensus quality: 206715 bases at least Q30
Consensus quality: 212351 bases at least Q20
Insert size: 217569; sum-of-contigs
Insert size: 165837; 6.2% error; agarose-fp
Quality coverage: 2.48x in Q20 bases; sum-of-contigs Quality
coverage: 3.95x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 6656: contig of 6656 bp in length
* 6657 6756: gap of 100 bp
* 6757 10255: contig of 3499 bp in length
* 10256 10355: gap of 100 bp
* 10356 13223: contig of 2868 bp in length
* 13224 13323: gap of 100 bp
* 13324 19353: contig of 6612 bp in length
* 19354 20035: gap of 100 bp
* 20036 22597: contig of 2562 bp in length
* 22598 22697: gap of 100 bp

* 22698 28569: contig of 5872 bp in length
* 28570 28669: gap of 100 bp
* 28670 37965: contig of 9296 bp in length
* 37966 38065: gap of 100 bp
* 38066 41140: contig of 3075 bp in length
* 41141 41240: gap of 100 bp
* 41241 45108: contig of 3868 bp in length
* 45109 45208: gap of 100 bp
* 45209 47730: contig of 2522 bp in length
* 47731 47830: gap of 100 bp
* 47831 51884: contig of 4054 bp in length
* 51885 51984: gap of 100 bp
* 51985 55857: contig of 3873 bp in length
* 55858 55957: gap of 100 bp
* 55959 60498: contig of 4541 bp in length
* 60499 60598: gap of 100 bp
* 60599 63641: contig of 3043 bp in length
* 63642 63741: gap of 100 bp
* 63742 72674: contig of 8933 bp in length
* 72675 72774: gap of 100 bp
* 72775 75602: contig of 2828 bp in length
* 75603 75702: gap of 100 bp
* 75703 78062: contig of 2360 bp in length
* 78063 78162: gap of 100 bp
* 78163 84485: contig of 6323 bp in length
* 84486 84585: gap of 100 bp
* 84586 87553: contig of 2968 bp in length
* 87554 87653: gap of 100 bp
* 87654 92487: contig of 4834 bp in length
* 92488 92587: gap of 100 bp
* 92589 96104: contig of 3517 bp in length
* 96105 96204: gap of 100 bp
* 96205 101479: contig of 5275 bp in length
* 101480 101579: gap of 100 bp
* 101580 104979: contig of 3400 bp in length
* 104980 105079: gap of 100 bp
* 105080 108003: contig of 2924 bp in length
* 108004 108103: gap of 100 bp
* 108104 113151: contig of 5048 bp in length
* 113152 113251: gap of 100 bp
* 113252 116585: contig of 3334 bp in length
* 116586 116685: gap of 100 bp
* 116686 119345: contig of 2661 bp in length
* 119347 119446: gap of 100 bp
* 119447 123125: contig of 3679 bp in length
* 123126 123225: gap of 100 bp
* 123226 131366: contig of 8041 bp in length
* 131367 131567: gap of 100 bp
* 131568 136193: contig of 4827 bp in length
* 136194 136293: gap of 100 bp
* 136294 139482: contig of 3189 bp in length
* 139483 139582: gap of 100 bp
* 139583 141875: contig of 2293 bp in length
* 141876 141975: gap of 100 bp
* 141976 144123: contig of 2147 bp in length
* 144123 144222: gap of 100 bp
* 144223 146968: contig of 2746 bp in length
* 146969 147068: gap of 100 bp
* 147069 149399: contig of 2331 bp in length
* 149400 149499: gap of 100 bp
* 149500 153963: contig of 4464 bp in length
* 153964 154063: gap of 100 bp
* 154064 156579: contig of 2516 bp in length
* 156580 156679: gap of 100 bp
* 156680 158928: contig of 2249 bp in length
* 158929 159028: gap of 100 bp
* 159029 164542: contig of 5514 bp in length
* 164543 164642: gap of 100 bp
* 164643 172054: contig of 7412 bp in length
* 172055 172154: gap of 100 bp
* 172155 175210: contig of 3056 bp in length
* 175211 175310: gap of 100 bp
* 175311 179379: contig of 4069 bp in length

* 179380 179479: gap of 100 bp
* 179480 182698: contig of 3219 bp in length
* 182699 182798: gap of 100 bp
* 182799 186779: contig of 3981 bp in length
* 186780 186879: gap of 100 bp
* 186880 195626: contig of 8747 bp in length
* 195627 195726: gap of 100 bp
* 195727 199396: contig of 3670 bp in length
* 199397 199496: gap of 100 bp
* 199497 206412: contig of 6916 bp in length
* 206413 206512: gap of 100 bp
* 206513 209619: contig of 3107 bp in length
* 209620 209719: gap of 100 bp
* 209720 219083: contig of 9364 bp in length
* 219084 219183: gap of 100 bp
* 219184 222469: contig of 3286 bp in length.
Location/Qualifiers
1. .222469
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKR-242K7"
/clone_lib="DanioKey"
1. .6656
/note="assembly fragment:01359
fragment chain:1"
misc_feature
6757. .10255
/note="assembly fragment:01385
fragment chain:1"
misc_feature
10356. .13223
/note="assembly fragment:00592
fragment chain:1"
misc_feature
13324. .19935
/note="assembly fragment:00237
fragment chain:1"
misc_feature
20036. .22597
/note="assembly fragment:00949
fragment chain:1"
misc_feature
22698. .38569
/note="assembly fragment:00615
fragment chain:1"
misc_feature
28670. .37965
/note="assembly fragment:01306
fragment chain:2"
misc_feature
38066. .41140
/note="assembly fragment:00487
fragment chain:2"
misc_feature
41241. .45108
/note="assembly fragment:00232
fragment chain:2"
misc_feature
45209. .47730
/note="assembly fragment:00935
fragment chain:2"
misc_feature
47831. .51864
/note="assembly fragment:01415
fragment chain:2"
misc_feature
51985. .55857
/note="assembly fragment:00878
fragment chain:3"
misc_feature
55958. .60498
/note="assembly fragment:01208
fragment chain:3"
misc_feature
60599. .63641
/note="assembly fragment:01266
fragment chain:3"
misc_feature
63742. .72674
/note="assembly fragment:01329
fragment chain:3"
misc_feature
72775. .75602
/note="assembly fragment:01183
fragment chain:4"
misc_feature
75703. .78062
/note="assembly fragment:01179

Query Match 8.7% Score 122.8; DB 2; Length 222469;
Best Local Similarity 88.7%; Pred. No. 1.6e-19;
Matches 133; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 206 GGTGCTTCCTGACTTCACCTTAGTCATGAGCCCATGCTTTCTTTATGATGCCGTT 265
DB 190751 GGTGATCTTCGTAAGTCACTGACAGTCAAGACCAAGCCCATGTTCTTCATGATGCCAT 190692
QY 266 TGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 325
DB 190691 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 190632
QY 326 AACCATTCAGGACGCTCCAGGTGATGCTG 355
DB 190631 AACCTTCAGGACGCTCCAGATGATGATG 190602

Search completed: December 16, 2003, 15:10:06
Job time : 5266.22 secs

ABK84798
 ID ABK84798 standard; cDNA, 220895 BP.
 XX
 AC ABK84798;
 XX
 DT 14-AUG-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #1369.
 XX
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 XX
 OS Homo sapiens.
 XX
 PN WO200228999-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 03-OCT-2001, 2001MO-US030821.
 XX
 PR 03-OCT-2000; 2000US-237189P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 DR WPI; 2002-435328/46.
 XX
 PT Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity.
 XX
 PS Claim 1; SEQ ID No 1369; 114pp; English.
 XX
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX
 SQ Sequence 220895 BP, 53760 A, 54808 C, 55831 G, 56496 T, 0 other;
 Query Match 20.0%; Score 282.6; DB 24; Length 220895;
 Best Local Similarity 74.6%; Pred. No. 3.6e-71;
 Matches 387; Conservative 0; Mismatches 119; Indels 13; Gaps 2;
 QY 554 ATATCAATATGATCTATGAGACCCACACACAGATATACATCTCAACAGAGGAAATA 613
 DB 169998 ATTCCAGTATATGCTATGAGACCCACCTGACAGATATGAGCCCACTCCGGATA 170057
 QY 614 TGGAACTCCACGAGAGATATGAGCCCAACGAGGGGATATGAGCCCACTTATGGG 673
 DB 170058 CGAGACCCCACTGACAGATATGAGCCCAACCTGAGAAATGAAAGCCCGCTGTGGG 170117
 QY 674 ATATGAGACCCCGCTGTGGGATATGAGTCCCACTTGGGGATATGAGTCCCACTTGG 733
 DB 170118 ATACAGAGCTTCACTGTGACATATGAGCCCACTTGTGATACGAGACCCCACTTGC 170177
 QY 734 GGGATATGAGATGCCACCTGGGGGATATGAGACCCCACTTGGGGATATGAGTCCCACT 793
 DB 170178 AGGATATGAGAGCCCACTTGAAGATATGAGCCCACTCTTGTGATATGAAACCCCACT 170237
 QY 794 TGGGGATATGATGTCGCCCACTTGGGGATATGAGACCCCACTTGAAGATATGAGCCCC 853
 DB 170238 TCTCGGATATGAGACCCCACTTGGATATGAGACCCCACTTGAAGAAATGAGGCC 170297
 QY 854 ACCAGCTGGAATGAAAGCCCTTACCCCTGTGATATGAACTCATCTGTGAAATATACAG 913
 DB 170298 GCTTCGGGATACAGAGCTTCACTGTGATATGAGACAGAGCTTCAGGAATCTACAGC 170357
 QY 914 TGCCTCTCAAGATCTATGACAGTCAAGAGAGACTTCTTCCCACTACCTCATCTTC 973
 DB 170358 AGCC-----CAGGCTCTGAAACAGAGCTTCTCTCTCTCTCTCTCTCTTC 170405
 QY 974 TTAGTCCATTACCACTTCTCAGAGTTAACTTGAAGCTCAACCAAG-AGCA 1032
 DB 170406 TAGGTCATTCTTAACCTTCAAGATGTAACCTTGAAGCTCAACCAAGAGGTA 170465
 QY 1033 CCTTAAACTGAAAGTCAAGTAAAGAAAGAACCCAGGT 1071
 DB 170466 CCTTAAATTTGAAGTCAAGATTAAGAGACGACTCAGGT 170504
 RESULT 4
 AL01229
 ID AL01229 standard; cDNA, 467 BP.
 XX
 AC AL01229;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DE Human reproductive system related antigen cDNA SEQ ID NO: 1230.
 XX
 KW Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200155320-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JUN-2001; 2001MO-US01339.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234227.
PR 21-SEP-2000; 2000US-0234227.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PR (HUMA-) HUMAN GENOME SCI INC.
PR PA
PR XX
PR PI Rosen CA, Barash SC, Ruben SM,
PR XX WPI; 2001-46570/50.
PR DR P-PSDB; AAM95259.
PR DR
PR XX
PR PT Isolated nucleic acid molecule encoding a reproductive system antigen
PR PT is used in preventing, treating or ameliorating a medical condition -
PR XX
PR PS Claim 1; SEQ ID NO 1230; 1297bp + Sequence Listing; English.
PR CC The present invention provides the protein and coding sequences of a

CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a coding sequence of the
CC invention.

XX Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;

Query Match 18.6%; Score 262.6; DB 22; Length 467;
Best Local Similarity 76.4%; Pred. No. 1,1e-66;
Matches 365; Conservative 0; Mismatches 100; Indels 13; Gaps 3;

QY 386 CAGAAAGAGAGTGCATGCAATTTGCCAATGATGTAAAGCTGCTGCTGCTGC 445
DB 2 CAGAAATGAGAGGATGATGATTTGCCAGTTGATGTGAAGGCTGCTGCTGCTGC 61
QY 446 CAGAGGAATTTCCCTTGGAGATTAATTAATGCTGCTGCTGCTGCTGCTGCTGCT 505
DB 62 CCGAGGATTTCCCTTGGAGATTAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 121
QY 506 TACTGTCGAGAGGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 565
DB 122 TACTGTCGAGAGGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 169
QY 566 GATCTATGAGACCCCAACAGATATACAGTCCACAGAGGGAATATGAACTCCACC 625
DB 170 TGTCTATGAGAGCCCACTGAGATATGAGCCCACTCCGGAATAGAGCCCAACC 229
QY 626 AGAAGATATGAGCCCAACAGAGATATGAGCCCACTGATGATATGAGCCCAACC 685
DB 230 TGCAGATATGAGCCCAACAGAGATATGAGCCCACTGATGATATGAGCCCAACC 289
QY 686 GCTGTGAGATATGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 745
DB 290 ACTGTGAGATATGAGCCCACTGCTGATGAGAGCCCACTGCTGATATGAGAGC 349
QY 746 CCCACCTGAGAGATATGAGCCCACTGCTGAGATATGAGTCCACCTGAGAGATATG 804
DB 350 CCCACCTGAGAGATATGAGCCCACTGCTGATGAGAGCCCACTGCTGATATGAG 409
QY 805 GTGCCCCACCTGAGAGATATGAGCCCACTGCTGAGATATGAGCCCACTGAGTGG 862
DB 410 GAGCCCCACCTGAGATATGAGCCCACTGCTGAGATATGAGCCCACTGAGTGG 467

RESULT 5
ABL96688
ID ABL96688 strand; cDNA; 467 BP.
XX
AC ABL96688;
XX
DT 21-JUN-2002 (first entry)
XX
DE Human testicular antigen encoding cDNA SEQ ID NO: 356.
XX
KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200155317-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216747.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226888.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229503.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.

PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-483232/52.
 XX
 XX Nucleic acids encoding 973 human testicular antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating testicular cancer
 XX
 XX
 PS Claim 1; SEQ ID NO 356; 766pp; English.
 XX
 XX The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and

CC gastrointestinal disorders, infections, and particularly cancer,
 CC especially testicular cancers. The present sequence is a cDNA of the
 CC invention.
 XX
 SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
 Query Match 18.6%; Score 262.6; DB 23; Length 467;
 Best Local Similarity 76.4%; Pred. No. 1.1e-66;
 Matches 365; Conservative 0; Mismatches 100; Indels 13; Gaps 3;
 QY 386 CAGGAAGAGGAGGCGCATGCAATTTGGCCACGATGATGTAAGCTGCTGTGCTGC 445
 DB 2 CAGGAATGAGGAGGCGCATGCAATTTGGCCACGATGATGTAAGCTGCTGTGCTGC 61
 QY 446 CAGGAATTCACCTTGAAGTGAATTTCTGCTGCACTTCAGACATATCAATAT 505
 DB 62 CCGAGATTCCTCACTTGAAGCTTGAATGACTGCTGATGGAATTTATGTAAT 121
 QY 506 TACTGTCCAGGGGCTGCACTGTGCTCTCCAGACACCTTGTCCAGCATATCAATG 565
 DB 122 TACTGGGAGGGGAAAT---ATGTCACTCCAGATGCTTTGAG-----TTAT 169
 QY 566 GATCTATGACCCCGACACGAGATATACATGTCACAGGGGATATGAACTCCACC 625
 DB 170 TGTCTATGAGCCCGACCTGAGATATGAGCCCGACCTCCGATACGAGCCCGACC 229
 QY 626 AGAAGATATGAGCCCGACGAGGATATGAGCCCGACCTATGAGATATGAGCCCGC 685
 DB 230 TGCAGATATGAGCCCGACGAGGATATGAGCCCGACCTATGAGATATGAGCCCGC 289
 QY 686 GCTGTGAGATATGAGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATGAGT 745
 DB 290 ACCTGTGAGATATGAGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATGAGT 349
 QY 746 CCGACCTGGGGATATGAGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATG 804
 DB 350 CCGACCTGGGGATATGAGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATG 409
 QY 805 GTGCCCGACCTGGGGATATGAGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATG 862
 DB 410 GAGCCCGACCTGGGGATATGAGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATG 467
 RESULT 6
 ABA67850
 ID ABA67850 standard; DNA; 436 BP.
 XX
 AC ABA67850;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DB Human foetal liver single exon nucleic acid probe #16155.
 XX
 KF Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX
 OS Homo sapiens.
 XX
 FN W0200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PP 30-JAN-2001; 2001WO-US00669.
 XX
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000US-0024263.
 XX
 XX (MOLB-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483447/52.
 DR Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human fetal liver -
 PT
 XX
 PS Claim 4; SEQ ID NO 16155; 639pp + sequence listing; English.
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human fetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 17.1%; Score 241.2; DB 22; Length 436;
 Best Local Similarity 73.2%; Pred. No. 2e-60;
 Matches 328; Conservative 0; Mismatches 108; Indels 12; Gaps 1;
 QY 562 TTGTGATCTATGAGACCCCAACAGATATACAGTCCAAACGAGGGAATATGAACTC 621
 DB 1 TTATGTCTATGAGACCCCAACAGATATGAGACCCCACTCCCGATACGAGCCC 60
 QY 622 CACGAGAAGATATGAGACCCCAACGAGGGAATATGAGACCCCACTATGGAATATGAG 681
 DB 61 CACTGCAAGATATGAGACCCCAACGAGGGAATATGAGACCCCACTATGGAATATGAG 120
 QY 682 CCCCCCTGTGGATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATG 741
 DB 121 CCTCACCTGTGGATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATG 180
 QY 742 GAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 801
 DB 181 GAGCCCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 240
 QY 802 ATGTGACCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 861
 DB 241 ATGAGACCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 300
 QY 862 GAAATGAGCCCTACCCCTGATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 921
 DB 301 GATACAGAGCTCAGCTGATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 356
 QY 922 ACAGATCTATGAGCTCAGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 981
 DB 357 -----CAGGCTCTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 408
 QY 982 ATTACCACTCTCAGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 1009
 DB 409 ATTCTTAACCTCTTAAGATGTAACCTT 436
 RESULT 7
 AAK42003
 ID AAK42003 standard; DNA; 436 BP.
 XX AAK42003;
 AC
 XX 06-NOV-2001 (first entry)
 DT
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 16560.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 XX macroarray; cancer; leukemia; lymphoma; myeloma; ss.
 OS Homo sapiens.
 XX

PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PP 30-JAN-2001; 2001WO-US000668.
 XX
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488900/53.
 DR
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 PS
 PS Example 4; SEQ ID NO: 16560; 658pp + Sequence listing; English.
 CC
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 CC
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 17.1%; Score 241.2; DB 22; Length 436;
 Best Local Similarity 73.2%; Pred. No. 2e-60;
 Matches 328; Conservative 0; Mismatches 108; Indels 12; Gaps 1;
 QY 562 TTGTGATCTATGAGACCCCAACAGATATACAGTCCAAACGAGGGAATATGAACTC 621
 DB 1 TTATGTCTATGAGACCCCAACAGATATGAGACCCCACTCCCGATACGAGCCC 60
 QY 622 CACGAGAAGATATGAGACCCCAACGAGGGAATATGAGACCCCACTATGGAATATGAG 681
 DB 61 CACTGCAAGATATGAGACCCCAACGAGGGAATATGAGACCCCACTATGGAATATGAG 120
 QY 682 CCCCCCTGTGGATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATG 741
 DB 121 CCTCACCTGTGGATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATG 180
 QY 742 GAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 801
 DB 181 GAGCCCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 240
 QY 802 ATGTGACCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 861
 DB 241 ATGAGACCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 300
 QY 862 GAAATGAGCCCTACCCCTGATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 921
 DB 301 GATACAGAGCTCAGCTGATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 356
 QY 922 ACAGATCTATGAGCTCAGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 981
 DB 357 -----CAGGCTCTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 408
 QY 982 ATTACCACTCTCAGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 1009
 DB 409 ATTCTTAACCTCTTAAGATGTAACCTT 436
 RESULT 8
 AAI48070

ID AAI48070 standard; DNA; 436 BP.
 AC AAI48070;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #16756 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KM genetic disorder; as.
 OS Homo sapiens.
 XX
 PN W0200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLR-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 XX
 PS analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 16756; 654bp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;

Query Match 17.1%; Score 241.2; DB 22; Length 436;
 Best Local Similarity 73.2%; Pred. No. 2e-60;
 Matches 328; Conservative 0; Mismatches 108; Indels 12; Gaps 1;

QY 562 TTGTGATCTATGACCCCAACAGATATACAGTCCACCAAGGGGAATATGAAATC 621
 DB 1 TTATTTGCTATGAGCCCACTGACAGATATGAGGCCCACTCCGATACAGCCC 60
 QY 622 CACGAGAAGATATGAGCCCAACAGGGGATATGAGGCCCACTATGGGATATGAG 681
 DB 61 CACTGAGAGATATGAGCCCAACCTAGGAATGAAAGGCCCTGTGGATACAGAG 120
 QY 682 CCCGCTGTGATATGAGTCCCACTGGGGATATGAGTCCCACTGGGGATATG 741
 DB 121 CTTCACTGTGATATGAGGCCCACTGTGATATGAGGCCCACTGTGAGATATG 180
 QY 742 GAATCCCACTGGGGATATGAGGCCCACTGGGGATATGAGTCCCACTGGGGAT 801
 DB 181 GAGCCCACTGTGATATGAGGCCCACTGTGATATGAGGCCCACTGTGATATG 240
 QY 802 ATGATGCCCCCACTGGGGATATGAGGCCCACTGTGATATGAGGCCCACTG 861
 DB 241 ATGAGGCCCACTGTGATATGAGGCCCACTGTGATATGAGGCCCACTGTG 300
 QY 862 GAAATGAGCCCTTACCCCTGTGATATGAGTCCATCTGTGATATGAGTCCCTCTC 921
 DB 301 GATATGAGCCCTTACCCCTGTGATATGAGTCCATCTGTGATATGAGTCCCTCTC 356

QY 922 ACAGATCTATGAGCTCAGCAGAGACTCTCTCCCACTATCTTATGAGTCC 981
 DB 357 -----CAGGCTCCGAAAACAGAGCTTCTTCCCTGCTCTCTTCAAGTCC 408
 QY 982 ATTACCACTTCTCAGAGTTAACTT 1009
 DB 409 ATTCTTAACTTCTCAGAGTTAACTT 436

RESULT 9
 ABS16034
 ID ABS16034 standard; DNA; 436 BP.
 XX
 AC ABS16034;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human genome-derived single exon probe ORF from lung SEQ ID No 16025.
 XX
 KW Human; de; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease; open reading frame; ORF.
 XX
 OS Homo sapiens.
 XX
 PN W0200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLR-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 WPI; 2002-114183/15.
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 XX
 PS measure gene expression in human lung samples -
 XX
 PS Claim 4; SEQ ID No 16025; 634bp; English.

The invention relates to a spatially-addressable set of single exon
 nucleic acid probes for measuring gene expression in a sample derived
 from human lung comprising single exon nucleic acid probes having one of
 12614 nucleic acid sequences mentioned in the specification, or their
 complements or the 12387 open reading frames derived from the 12614
 probes. Also included are a microarray comprising the novel set of
 probes; the novel set of probes which hybridise at high stringency to a
 nucleic acid expressed in the human lung; measuring gene expression in a
 sample derived from human lung, comprising (a) contacting the array with
 a collection of detectably labeled nucleic acids derived from human lung
 mRNA, and (b) measuring the label detectably bound to each probe of
 the array; identifying exons in a eukaryotic genome, comprising
 (a) algorithmically predicting at least one exon from genomic sequences
 of the eukaryote; and (b) detecting specific hybridisation of detectably
 labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 having a fragment identical to the predicted exon, the probe is included

in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarray having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Rudlak syndrome, sarcoidosis, pulmonary haemoblastosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a single exon probe open reading frame of the invention.

Note: the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIP0 at ftp.wipo.int/pub/published_pct_sequences.

Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;

Query Match 17.1%; Score 241.2; DB 24; Length 436;
Best Local Similarity 73.2%; Pred. No. 2e-60;
Matches 328; Conservative 0; Mismatches 108; Indels 12; Gaps 1;

```
QY 562 TTGTATCTATGAGACCCCAACCAAGATATACATCCCAAGGAGATATGAACTC 621
DB 1 TTATTTCTATGAGGCCCACTGCAAGATATGAGCCCACTCCGATACGAGCCC 60
QY 622 CACCAAGAGATATGAGGCCCAACCAAGGAGATATGAGGCCCACTATGAGATGAG 681
DB 61 CACCTGAGAGATATGAGGCCCAACCAAGGAGATATGAGGCCCACTGAGATACAGAG 120
QY 682 CCCCCTGTGAGATATGAGTCCCACTGGGAGATATGAGTCCCACTGGGAGATATG 741
DB 121 CCACTGAGATATGAGGCCCACTGAGATACGAGGCCCACTGAGATATG 180
QY 742 GAGTCCCACTGGGAGATATGAGGCCCACTGGGAGATATGAGTCCCACTGGGAGAT 801
DB 181 GAGCCCACTGAGATATGAGGCCCACTGAGATATGAGGCCCACTGAGAT 240
QY 802 ATGGTCCCACTGGGAGATATGAGGCCCACTGAGATATGAGGCCCACTGAGAT 861
DB 241 ATGGAGCCCACTGAGATATGAGGCCCACTGAGATATGAGGCCCACTGAGAT 300
QY 862 GAATGAGCCCACTGAGATATGAGGCCCACTGAGATATGAGGCCCACTGAGAT 921
DB 301 GATGAGCCCACTGAGATATGAGGCCCACTGAGATATGAGGCCCACTGAGAT 356
QY 922 ACGAGTATGAGCACTGAGAGGAGACTTCTTCCCACTACTCTATCTTTAGATCC 981
DB 357 -----CAGGCTCTCGAAGAACGAGGCTTCTTCCCTGCTCTCTTCAAGTCC 408
QY 982 ATTACACCTCTGAGATTAACCTT 1009
DB 409 ATTCTTAACCTTCAAGATTAACCTT 436
```

RESULT 10

AA04882 standard; DNA; 7099 BP.

AA04882;

21-NOV-2001 (first entry)

Human reproductive system related antigen DNA SEQ ID NO: 7570.

```
XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; da.
XX Homo sapiens.
XX WO20015320-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01339.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.
```

PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0234984.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237040.
 PR 02-OCT-2000; 2000US-0237049.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246603.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251866.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HOMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM,
 DR WPI, 2001-465570/50.
 XX
 PT Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 XX
 PS Disclosure; SEQ ID NO 7570; 1297bp + Sequence Listing; English.
 CC
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a genomic sequence encoding a
 CC protein of the invention.
 XX
 SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;

 Query Match 15.6%; Score 220; DB 22; Length 7099;
 Best Local Similarity 75.3%; Pred. No. 1.5e-53;
 Matches 274; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

 QY 554 ATATCCAAATTGTATCTATGAGACCCCAACCAAGATATACAGTCCCAAGGGGAAATA 613
 DB 6718 ATTCGAGTATATCTATGAGACCCCACTGCGAGATATGAGACCCCACTCCGGATA 6777
 QY 614 TGAACCTCACAGAGATATGAGAGCCCAACGAGGGGATATGAGACCCCACTATGGG 673
 DB 6778 CGAGGCCCACTGAGATATGAGAGCCCAACCTTGAGAAATAGAGCCCGCTGTGGG 6837
 QY 674 ATATGAGCCCGGCTGTGAGATATGAGTCCCACTGGGGATATGAGTCCCACTGG 733
 DB 6838 ATACAGAGCTCACTGTGAGATATGAGGCCCACTCTTGATATACGAGGCCCACTGG 6897
 QY 734 GGGATATGAGTCCCACTGGGGATATGAGAGCCCACTGGGGGATATGAGTCCCACT 793
 DB 6898 AGGATATGAGCCCACTCTGAGATATGAGAGCCCACTCTTGATATGAGAGCCCACT 6957
 QY 794 TGGGAGATATGAGTCCCACTGGGGATATGAGAGCCCACTGAGAGATATGAGAGCC 853
 DB 6958 TCTCGATATGAGAGCCCACTCTGAGATATGAGAGCCCACTGAGAGATATGAGAGCC 7017
 QY 854 ACCAGCTGGAATGAGAGCCCACTCTGAGATATGAGTCCCACTGAGAGATATGAGC 913
 DB 7018 GCGTGGGGATATGAGAGCCCACTCTGAGATATGAGAGCCCACTGAGAGATATGAGC 7077
 QY 914 TGCC 917
 DB 7078 AGCC 7081

 RESULT 11
 ABL97776
 ID ABL97776 standard; DNA; 7099 BP.
 XX
 AC ABL97776;
 XX
 DT 21-JUN-2002 (first entry)
 XX
 DB Human testicular antigen encoding DNA fragment SEQ ID NO: 2428.
 XX
 KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
 KW reproductive system disorder; urinary system disorder; gene therapy;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disease; infection; cytotoxic; gene; ds.
 XX
 OS Homo sapiens.
 OS
 XX
 FN W0200155317-A2.

XX 02-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US01329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-024475.
PR 08-NOV-2000; 2000US-024476.
PR 08-NOV-2000; 2000US-024477.
PR 08-NOV-2000; 2000US-024478.
PR 08-NOV-2000; 2000US-024479.
PR 08-NOV-2000; 2000US-0244823.
PR 08-NOV-2000; 2000US-0244824.
PR 08-NOV-2000; 2000US-0244825.
PR 08-NOV-2000; 2000US-0244826.
PR 08-NOV-2000; 2000US-0244827.
PR 08-NOV-2000; 2000US-0244828.
PR 08-NOV-2000; 2000US-0244829.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254907.
PR 05-JAN-2001; 2001US-02559678.

(HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI Rosen CA, Barash SC, Ruben SW,
XX

DR WPI; 2001-483232/52.
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
PT useful for preventing, diagnosing and/or treating testicular cancer
XX
XX Disclosure; SEQ ID NO 2428; 766bp; English.
XX
CC The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancer. The present sequence is a DNA encoding a
CC protein fragment of the invention.
XX
SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
Query Match 15.6%; Score 220; DB 23; Length 7099;
Best Local Similarity 75.3%; Pred. No. 1.5e-53;
Matches 274; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 554 ATATCAATTGTGATCTATGAGACCCCAACAGATATACAGTCCACAGGGAATA 613
DB 6718 ATCCCAATTATGTCTATGAGCCCACTGAGATATGAGCCCACTCCGATTA 6777
QY 614 TGGAACTCCACAGAGATATGAGCCCAACAGGGAATATGAGCCCACTTATGG 673
DB 6778 CGAGGCCCACTGAGATATGAGCCCAACAGGGAATATGAGCCCACTTATGG 6837
QY 674 ATATGAGCCCGCTGAGATATGAGTCCCACTGAGGGAATATGAGCCCACTGG 733
DB 6838 ATACAGAGCCCTGAGATATGAGCCCACTTGTGATATGAGCCCACTTATGG 6887
QY 734 GGGATATGAGTCCCACTGAGGGAATATGAGCCCACTGAGGGAATATGAGTCCCA 793
DB 6898 AGATATGAGAGCCCACTTATGAGATATGAGCCCACTTGTGATATGAGCCCA 6957
QY 794 TGGGGAATATGAGTCCCACTGAGGGAATATGAGCCCACTGAGATATGAGCC 853
DB 6958 TCTCGATATGAGAGCCCACTTGTGATATGAGCCCACTTGTGATATGAGAG 7017
QY 854 ACCAGTGAATATGAGAGCCCACTGAGTATGAGTCCATCTGAGATATGAG 913
DB 7018 GCTTCGAGATATGAGAGCTTCACTGATATGAGAGCCCACTTCAAGATCTA 7077
QY 914 TGCC 917
DB 7078 AGCC 7081
RESULT 12
AAL01355
ID AAL01355 standard; cDNA; 321 BP.
XX
XX AAL01355;
XX
XX 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen cDNA SEQ ID NO: 1356.
XX
XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX W0200155320-A2.
XX
XX
XX 02-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001MO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 15-MAR-2000; 2000US-0190874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216477.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0225799.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226682.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232402.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234977.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PR Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-465570/50.
 DR P-PSDB; AAM95385.
 XX Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 XX Claim 1; SEQ ID NO 1356; 1297bp + Sequence Listing; English.

XX The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a coding sequence of the
 CC invention.
 XX
 SQ Sequence 321 BP; 83 A; 75 C; 70 G; 89 T; 4 other;
 Query Match 13.8%; Score 195.4; DB 22; Length 321;
 Best Local Similarity 77.4%; Pred. No. Se-47; Mismatches 69; Indels 1; Gaps 1;
 Matches 246; Conservative 2;
 QY 64 GCCGTCGAGGCGCCCTCATCCCTCGCGAAGAGTCTTGAAGCAGTGTGAGATGCG 123
 DB 1 GCCGCGCGAGACCTCATCCCTAAGCGAAGTCTTGAAGCGGCTCGAATGCG 60
 QY 124 ACCTTGCTCTCAAGAAACGATGGAATCTATCTTTAAGCACAAGAAAGANA 183
 DB 61 AGCTCTCTCCACAGGAGATCAGAAAGCTCAAAATGCTTTAGTGTAGAAAGACAGGA 120
 QY 184 CGTTGTTCTCACTTATACCGGGGCTTCGTGACTTCACTTAAGTGAATGACCCCA 243
 DB 121 CATTTGTTCTCACTTATACCGGGGCTTCGTGACTTCACTTCACTTCACTTCACTTCA 180
 QY 244 TGCCTTCTTTAAGATGCGGTTGGCTGATGAGTCACTGACCAATGAACCAATTT 303
 DB 181 TGTGCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 QY 304 TTGCCCCCACTTCAATTAAGAACCAATTCAGGACCTTCAGGTGCTGGG-AAGGA 362
 DB 241 TTGCTGCAAACTTCAATTAAGGAACTATTCAGGACCTTCAGGTGCTGGG-AAGGA 300
 QY 363 CAAGCTTTTAAAGTAA 380
 DB 301 CAAGCTTTTAAAGTAA 318
 Db
 RESULT 13
 ABL96808
 ID ABL96808 standard; cDNA, 321 BP.
 XX ABL96808;
 AC
 XX
 DT 21-JUN-2002 (first entry)
 XX
 XX
 DE Human testicular antigen encoding cDNA SEQ ID NO: 476.
 XX
 XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
 KM reproductive system disorder; urinary system disorder; gene therapy;
 KM cardiovascular disorder; respiratory disorder; neurological disorder;
 KM gastrointestinal disease; infection; cytostatic; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX
 PN MO200155317-A2.
 PD
 XX
 XX 02-AUG-2001.
 PF 17-JAN-2001; 2001WO-US01329.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.

```
PR 07-JUL-2000; 2000US-021680.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226682.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483232/52.
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
PT useful for preventing, diagnosing and/or treating testicular cancer -
XX
XX Claim 1; SEQ ID NO 476, 766pp; English.
XX
XX The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a cDNA of the
CC invention.
XX
XX Sequence 321 BP; 83 A; 75 C; 70 G; 89 T; 4 other;
```

Query Match 13.8%; Score 195.4; DB 23; Length 321;
 Best Local Similarity 77.4%; Pred. No. 5e-47;
 Matches 246; Conservative 2; Mismatches 69; Indels 1; Gaps 1;

QY 64 GCGCGTGTGGGCGCTCATCCCTCTGGCGAAAGTCTTGAAGCAGTGTAGATGTG 123
 DB 1 GCGCGCGGAGCGCTCATCCCTTACGGTGAAGTCTTGAAGCGGTCTCCGAATGTG 60
 QY 124 ACGTCTTCCACAGAAACAGTGAATCCATCTCTTATGCGAAAGAAAGAA 183
 DB 61 AGCTCTCTCCACAGCAATCAAGAGCTCAAAATCTTATGTGTAGAAACAGAA 120
 QY 184 CGTGTCTCACTTCAATCCGGGTGTCTTCTGCACTTCACTTCAATGACCCCA 243
 DB 121 CATGTCTCACTTCAATCCGGGTGTCTTCTCAATCTTCAATGACCCCA 180
 QY 244 TCGTCTTCTTATGATGCGGTGTGGCTGATGATGATGATGATGATGATGAT 303
 DB 181 TGTGTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 QY 304 TTGCCCCCACTCAATTAAGAAACATTCAGGCACTCCAGTGTGTGTGG-AGA 362
 DB 241 TTGCTCAAACTTCATTAAGGAACTATTCAGGCACTCAATGAGCTGGAAAGA 300
 QY 363 CAAGCTGTTTTTAAGTTA 380
 DB 301 CAAGCTACTTTTAANTTA 318

RESULT 14
 ABS70430
 ID ABS70430 standard; cDNA; 1885 BP.
 AC ABS70430;
 DT 27-NOV-2002 (first entry)
 XX
 DE Human bone remodelling gene #87.
 XX
 KW Bone remodelling; osteoporosis; human; gene; 88.
 XX
 OS Homo sapiens.
 XX
 PN US6426186-B1.
 PD 30-JUL-2002.
 XX
 PF 18-JAN-2000; 2000US-0484970.
 XX
 PR 18-JAN-2000; 2000US-0484970.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Jones KA, Volkmutz W, Walker MG;
 XX
 DR WPI; 2002-673014/72.
 XX
 PT A combination of polynucleotides which are co-expressed with genes
 PT known to be involved in bone remodeling and osteoporosis are useful in
 PT an array for the diagnosis of bone remodeling and osteoporosis
 PT associated disorders -
 XX
 PS Claim 1; Column 247-250; 206pp; English.
 XX
 CC The invention relates to a combination comprising a number of
 CC substantially purified and isolated polynucleotides which are
 CC co-expressed with genes known to be involved in bone remodeling and
 CC osteoporosis. The invention is used to diagnose disorders associated
 CC with bone remodeling or osteoporosis. ABS70344-ABS70512 represent
 CC human bone remodelling genes of the invention.
 XX
 SQ Sequence 1885 BP; 384 A; 610 C; 476 G; 415 T; 0 other;

Query Match 8.6%; Score 122; DB 24; Length 1885;
 Best Local Similarity 59.4%; Pred. No. 5e-25;
 Matches 247; Conservative 0; Mismatches 160; Indels 9; Gaps 2;

QY 36 ATGCAATGAAACCAAGCCACACGAGAGCGGTGTGTGGGCGCTCATCCCTCTGGCGAA 95
 DB 61 ATGGGCTTCACCAAAATCACTTCGAGGGCGGAGATGATGATGATGATGATGATGAT 117
 QY 96 AGTGTCTTGAAGCAATGTGAGATGTGAGACCTGTCTCTCTTACAGAAACAGTGTATCC 155
 DB 118 AGCATCTTAATGTCTTATGATCACTGTGAATCTCACTTCAATGATGAAGAGTGGCA 177
 QY 156 TATCTCTTAAATGCAAAAGAAAGAAAGTGTGTTCTCACTTCAATCCGGGTGTCTTC 215
 DB 178 GAAGCTTCAAAAGGACCAAGAAAGGCACTGTCTTCACTTCACTTCACTTCACTTCTT 237
 QY 216 GTGACTTCACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 275
 DB 238 CTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 291
 QY 276 AGTCACTGACCATTTGAACCAACCAATTTTGGCCCCCACTAATTAAGAAACCATTCAG 335
 DB 222 AAGACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 351
 QY 336 GCACTCCAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 395
 DB 352 GCGAAGGGGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 411
 QY 396 GGTGCAATGCAATTTGCCCACTGATGATGATGATGATGATGATGATGATGATGATG 451
 DB 412 GCGGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 467

RESULT 15
 AAF22363
 ID AAF22363 standard; cDNA; 1915 BP.
 AC AAF22363;
 DT 26-MAR-2001 (first entry)
 XX
 DE Human secreted protein gene 48 SEQ ID NO:58.
 XX
 KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytoskeletal; cardiac; vasotropic;
 KW cerebroprotective; noctropic; neuroprotective; antibacterial; vitruide;
 KW fungicide; ophthalmological; vulnary; gene therapy; neoplasm;
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; angiodenesis; nervous system disorder; infection;
 KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;
 KW skin aging; food additive; preservative; 88.
 XX
 OS Homo sapiens.
 XX
 PN WO200061748-A1.
 PD 19-OCT-2000.
 XX
 PF 06-APR-2000; 2000WO-US08982.
 XX
 PR 09-APR-1999; 99US-0128696.
 XX
 PR 14-JAN-2000; 2000US-0176069.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis G;
 XX
 DR WPI; 2000-638566/61.
 DR P-PSDB; AAB63096.
 XX
 PT New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and

PT used as food additives or preservatives -
XX
PS Claim 1, Page 429-430, 480pp; English.

CC AAF22316 to AAF22363 encode the human secreted proteins given in AAB63049
CC to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins
CC and polypeptides homologous to them. Human secreted proteins have
CC activities based on the tissues and cells the genes are expressed in.
CC Examples of activities include: immunosuppressive; antiarthritic;
CC antirheumatic; antiproliferative; cyrostatic; cardiant; vasotropic;
CC cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
CC fungicide; ophthalmological; and vulnerary. The polynucleotides and
CC proteins can be used to prevent, treat or ameliorate a medical condition
CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
CC sheep. They are also used in diagnosing a pathological condition or
CC susceptibility to a pathological condition. Disorders which are diagnosed
CC or treated include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. The polypeptides can also be
CC used to aid wound healing and epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities, fat content,
CC lipid, protein, carbohydrate, vitamins, minerals, cofactors and other
CC nutritional components. AAF22307 to AAF22315 and AAB63048 represent
CC sequences used in the exemplification of the present invention.

XX
SQ Sequence 1915 BP; 401 A; 613 C; 481 G; 413 T; 7 other;

Query Match 8.6%; Score 122; DB 21; Length 1915;
Best Local Similarity 59.4%; Pred. No. 5e-25;
Matches 247; Conservative 0; Mismatches 160; Indels 9; Gaps 2;

QY 36 ATGGAGTGAACCAAGCAGCCGAGCCGCTGCTGAGCCCTCATCCCTTGGCGAA 95
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
59 ATGGGCTCAACAAATCACTCGAGGCGCGGAGTATGTCATATACAC--CGAG 115
QY 96 AGTGCTTGAAGCAGTGTAGAGATGTGACCTGTGCTTCTACAGAAACAGTGAATCC 155
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
116 AGCATCTTAATGTCTTATGATCACTGGAATCAATCATATGACATGAAGAAGCTGCCA 175
QY 156 TATCTTTAATGCAAAAGAAAGAAAGTTGTTCTACTTCATACCGGAGTGTCTTC 215
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
176 GAAGCTTCAAGGGAACAAGAAAGGCACTGTCTACCTTACCCCTTACCGGATCATCTTT 235
QY 216 GTGACTTCACTTATGATGACCCCATGCTTTCTTTATGATGCCGTTGGCTGATG 275
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
236 CTGTC-----CAAGGCGAAGATGCAATGCAATGCTTCATGATGCCATTTATCTCATG 289
QY 276 AGTGACTGACACATTTGAACAACAAATTTTGGCCCCCACTACATTTAAGAAACCATTCAG 335
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
290 AAAGACTGTGATATCAAGACGCCGTATTTGGTGCAACATCAACAGGAACAGTGAAG 349
QY 336 GCAAGCTCAGGTGTGCTGGGAAGACAGCTGTTTAAATTATCCTTCAGAAAGAA 395
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 GCGAAGCGGAGGAGGTGGTCTGCTTCTTACAGTTGACTTTACCGGCAAGG 409
QY 396 GTGCGATCGAATTTGCCCAACTGATGTGTAAGTGTGCTGCTGCTGCCAGAG 451
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
410 GCGCGCATGTGATTCGAGACGCGAGTCTCCAGGTGATCTCAAGCTCCAGAGG 465

Search completed: December 16, 2003, 11:00:45
Job time : 382.081 secs

TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
FILE REFERENCE: 16412-100018
CURRENT APPLICATION NUMBER: US/09/298,568
CURRENT FILING DATE: 1999-04-21
EARLIER APPLICATION NUMBER: US 60/109,422
EARLIER FILING DATE: 1998-11-19
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 3489
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1

Query Match 4.2%; Score 59.8; DB 4; Length 3489;
Best Local Similarity 46.9%; Pred. No. 4.9e-08;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 514 CAGGGGCTGCAGTGTCTCTCAGACAGACCTTGTCCAGCATTCATTTGTATCTATG 573
DB 2185 CAGCAGCAGCAGATGAGCAGCAGCAGATGACGACGACGATGACGACGAG 2244
QY 574 GACCCCAACACACAGATATACGTCCAAACGAGGATATGAACTCCACAGAGAT 633
DB 2245 CAGCAGATGAAACAGACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2304
QY 634 ATGAGCCCAACAGGAGGATATGAGCCCACTTATGATATGAGCCCGCTGTG 693
DB 2305 TTACAGAGCAGCAGAGAGATGATGATGATGATGATGATGATGATGATGATG 2364
QY 694 GATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 753
DB 2365 GAGTTAGAGAGCAGCAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAG 2424
QY 754 GGGATATGAGACCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACT 813
DB 2425 CAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2484
QY 814 CTGGGGATATGAGACCCCACTGAGGATATGAGCCCACTGAGGATATGAGCC 873
DB 2485 GAGCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2544
QY 874 TACCCCTGCATATGAGCTTCATCTGTGAATAACAG 912
DB 2545 CAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2583

RESULT 5
US-09-410-399-1
Sequence 1, Application US/09410399
Patent No. 6482587
GENERAL INFORMATION:
APPLICANT: Robertson, Erle S.
APPLICANT: Cotter, Murray A.
TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
FILE REFERENCE: US-03778
CURRENT APPLICATION NUMBER: US/09/410,399
CURRENT FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 3489
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-1

Query Match 4.2%; Score 59.8; DB 4; Length 3489;
Best Local Similarity 46.9%; Pred. No. 4.9e-08;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;
QY 514 CAGGGGCTGCAGTGTCTCTCAGACAGACCTTGTCCAGCATTCATTTGTATCTATG 573

DB 2185 CAGCAGCAGCAGATGAGCAGCAGCAGATGACGACGACGATGACGACGAG 2244
QY 574 GACCCCAACACACAGATATACGTCCAAACGAGGATATGAACTCCACAGAGAT 633
DB 2245 CAGCAGATGAAACAGACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2304
QY 634 ATGAGCCCAACAGGAGGATATGAGCCCACTTATGATATGAGCCCGCTGTG 693
DB 2305 TTACAGAGCAGCAGAGAGATGATGATGATGATGATGATGATGATGATGATG 2364
QY 694 GATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 753
DB 2365 GAGTTAGAGAGCAGCAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAG 2424
QY 754 GGGATATGAGACCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACT 813
DB 2425 CAGAGATGAGAGAGAGAGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAG 2484
QY 814 CTGGGGATATGAGACCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACT 873
DB 2485 GAGCAGAGATGAGAGAGAGAGAGAGAGAGATGATGAGAGAGAGAGAGAGAG 2544
QY 874 TACCCCTGCATATGAGCTTCATCTGTGAATAACAG 912
DB 2545 CAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2583

RESULT 6
US-08-770-379-20/c
Sequence 20, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Ruoso, James J.
APPLICANT: Radelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSER: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULAR TYPE: DNA (genomic)
US-08-770-379-20

Query Match 4.2%; Score 59.8; DB 2; Length 32207;

Best Local Similarity 46.9%; Pred. No. 1.9e-07;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 514 CAGGGGCTGCAAGTGTCTCTCTCAAGACACCTTGTCCAGATATCCAAATGTGATCTATG 573
DB 19812 CAGCAGCAGCAGATGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAG 19753

QY 574 GACCCCAACCAACAGATATACGTCCCAACAGGGGAAATATGAACTCCACCAAGAGAT 633
DB 19752 CAGCAGATGAAACAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19693

QY 634 ATGAGCCCAACCAAGGGGATATGAGCCCACTATGGAATATGAGCCCGCTGTG 693
DB 19692 TTAGAGAGCAGAGCAGAGGTTAGAGATCAGAGCAGAGGTTAGAGAGCAGAGCAG 19633

QY 694 GATATGAGTCCCACTGAGGGGATATGAGTCCCACTGAGGGATATGAGTCCCACTG 753
DB 19632 GAGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAGAGAGAGAGGTTAGAGAGCAGAG 19573

QY 754 GGGGATATGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGATATGAGTCCCACTG 813
DB 19572 CAGGAGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAG 19513

QY 814 CTGGGGATATGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGATATGAGTCCCACTG 873
DB 19512 GAGCAGAGGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAGAGAGGTTAGAGAGCAG 19453

QY 874 TACCCCTGATATGAGCTTCATCTGTGGAATATGAG 912
DB 19452 CAGGAGTGTGAAGAGCAGAGCAGAGCAGAGGAGAGCAG 19414

RESULT 7

US-08-757-669A-20/c
Sequence 20, Application US/08757669A
Patent No. 6183751

GENERAL INFORMATION:

APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNICORE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757,669A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-F
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20

Query Match 4.2%; Score 59.8; DB 3; Length 32207;
Best Local Similarity 46.9%; Pred. No. 1.9e-07;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 514 CAGGGGCTGCAAGTGTCTCTCTCAAGACACCTTGTCCAGATATCCAAATGTGATCTATG 573
DB 19812 CAGCAGCAGCAGATGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAG 19753

QY 574 GACCCCAACCAACAGATATACGTCCCAACAGGGGAAATATGAACTCCACCAAGAGAT 633
DB 19752 CAGCAGATGAAACAGAGCAGAGCAGAGGTTAGAGATCAGAGCAGAGGTTAGAGAGCAGAG 19693

QY 634 ATGAGCCCAACCAAGGGGATATGAGCCCACTTATGGAATATGAGCCCGCTGTG 693
DB 19692 TTAGAGAGCAGAGCAGAGGTTAGAGATCAGAGCAGAGGTTAGAGAGCAGAGCAG 19633

QY 694 GATATGAGTCCCACTGAGGGGATATGAGTCCCACTGAGGGATATGAGTCCCACTG 753
DB 19632 GAGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAGAGAGAGGTTAGAGAGCAGAG 19573

QY 754 GGGGATATGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGATATGAGTCCCACTG 813
DB 19572 CAGGAGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAG 19513

QY 814 CTGGGGATATGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGATATGAGTCCCACTG 873
DB 19512 GAGCAGAGGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAGAGAGGTTAGAGAGCAG 19453

QY 874 TACCCCTGATATGAGCTTCATCTGTGGAATATGAG 912
DB 19452 CAGGAGTGTGAAGAGCAGAGCAGAGCAGAGGAGAGCAG 19414

RESULT 8

US-09-230-371A-20/c
Sequence 20, Application US/09230371A
Patent No. 6348586

GENERAL INFORMATION:

APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNICORE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: 45185-G-PCT-US
CURRENT FILING DATE: US/09/230,371A
PRIOR FILING DATE: 1999-11-17
PRIOR APPLICATION NUMBER: PCT/US97/13346
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 20
LENGTH: 32207
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

Query Match 4.2%; Score 59.8; DB 4; Length 32207;
Best Local Similarity 46.9%; Pred. No. 1.9e-07;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 514 CAGGGGCTGCAAGTGTCTCTCTCAAGACACCTTGTCCAGATATCCAAATGTGATCTATG 573
DB 19812 CAGCAGCAGCAGATGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAG 19753

QY 574 GACCCCAACCAACAGATATACGTCCCAACAGGGGAAATATGAACTCCACCAAGAGAT 633
DB 19752 CAGCAGATGAAACAGAGCAGAGCAGAGGTTAGAGATCAGAGCAGAGGTTAGAGAGCAGAG 19693

QY 634 ATGAGGCCCAACGAGGGGATATGAGCCCACTATGAGTATGAGCCCGCTGTG 693
DB 19692 TTATAGAGGACGAGGAGTTAGAGATGAGAGCAGAGGTTAGAGAGCAGAGCAG 19633
QY 694 GATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 753
DB 19632 GAGTTAGAGGACGAGGAGGAGTTAGAGAGCAGAGGAGTTAGAGAGCAGAG 19573
QY 754 GGGATATGAGCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACT 813
DB 19572 CAGAGTTAGAGGACGAGGAGGAGTTAGAGAGCAGAGGAGTTAGAGAGCAG 19513
QY 814 CTGGGGATATGAGCCCACTGAGGATATGAGCCCACTGAGGATATGAGCCCACT 873
DB 19512 GAGAGGAGTTAGAGGACGAGGAGGAGTTAGAGAGCAGAGGAGTTAGAGAGCAG 19453
QY 874 TACCCCTGATGAGGCTCATCTGCTGGAATGAG 912
DB 19452 CAGAGGTTGAGAGGACGAGGAGCAGAGCAGAGCAGAGCAG 19414

RESULT 9

US-08-917-320-18

Sequence 18, Application US/08917320

Patent No. 5824508

GENERAL INFORMATION:

APPLICANT: Spaete, Richard and Jackman, Winthrop, T.

TITLE OF INVENTION: No. 5824508 Splicing Variants of gp350/220

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESSES:

STREET: 5 Palo Alto Square

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/917,320

CLASSIFICATION: 435

FILING DATE: 25-AUG-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/229,291

FILING DATE: April 18, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Luann Caert

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: AVIR-003/00US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-843-5163

TELEFAX: 415-857-0663

TELEX: 380816 COOLEYPA

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 3833 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1014..3734

US-08-917-320-18

Query Match 3.9%; Score 54.8; DB 1; Length 3833;
Best Local Similarity 48.7%; Pred. No. 1.8e-06;
Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 551 AGCATATCCATTGTGATCTATGACCCCAACCAAGGATATACATGTCACCAAGGGA 610
DB 2532 ACCAGCTCACTCAACGAGTACTACCCCAACCCCAATGCCACAGCCCACTCAAGCA 2591
QY 611 ATATGAACTCCACCAAGAGATATGAGCCCAACGAGGAGTATGAGCCCACTAT 670
DB 2592 GTGACTTACCCCAACCCCAATGCCACCAAGCCCACTCAAGTACTACCCCAACCCA 2651
QY 671 GGGATATGAGCCCGCTGTGGATATGAGTCCCACTGAGGAGTATGAGTCCCACT 730
DB 2652 AATGCCACGAGCCCACTTGGAAAAACAAGTCTTACCTCAGACAGTACTACCCCACT 711
QY 731 TGGGGATATGAGTCCCACTGAGGATATGAGCCCACTGAGGAGTATGAGTCCCACT 790
DB 2712 CCAATGCCACAGCCCACTTGGAAAAACAAGCCCACTCAGACAGTACTACCCCA 2771
QY 791 ACCTGGGAGTATGATGTCCTCCCACTGAGGATATGAGCCCACTGAGAGATATGAGC 850
DB 2772 ACCCAATGCCACAGCCCACTTGGAAAAACAAGCCCACTCAGACAGTACTACCC 2831
QY 851 CCAACC 856
DB 2832 CCAACC 2837

RESULT 10

PCT-US95-04611A-18

Sequence 18, Application PC/TUS9504611A

GENERAL INFORMATION:

APPLICANT: Spaete, Richard and Jackman, Winthrop, T.

TITLE OF INVENTION: Non Splicing Variants of gp350/220

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESSES:

STREET: 5 Palo Alto Square

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04611A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/229,291

FILING DATE: April 18, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Luann Caert

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: AVIR-003/00US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-843-5163

TELEFAX: 415-857-0663

TELEX: 380816 COOLEYPA

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 3833 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1014..3734

PCT-US95-04611A-18

Query Match 3.9%; Score 54.8; DB 5; Length 3833;

Best Local Similarity 48.7%; Pred. No. 1.8e-06;
Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 551 AGCATATCCAAATTGATCTATGAGCCCAACCAAGATATACAGTCCACCAAGGGA 610
DB 2532 ACCAGCTCACTCAACGATGACTACCCCAACCCCAATGCAACCAAGCCCAACCA 2591
QY 611 ATATGAACTCCACCAAGAGATATGAGCCCAACCAAGGAGATATGAGCCCACTAT 670
DB 2532 GTGACTACCCCAACCCCAATGCAACGAGCCCAACCAAGTACTACCCCAACCA 2651
QY 671 GGGATATGAGAGCCCGCTGTGGATATGAGTCCCACTGGGGATATGAGTCCCA 730
DB 2652 AATGCCACACGCCCCCACTTGGGAAAAACAAGTCTACTGACAGTGACTACCCCAAC 2711
QY 731 TGGGGATATGAGTCCCACTGGGGATATGAGGCCCCCACTGGGGATATGAGTCC 790
DB 2712 CCAATGCCACACGCCCCCACTTGGGAAAAACAAGTCTACTGACAGTGACTACCCCA 2771
QY 791 ACTGGGGATATGAGTCCCACTGGGGATATGAGGCCCCCACTGGAGATATGAGAC 850
DB 2772 ACCCAATGCCACACGCCCCCACTTGGGAAAAACAAGTCTACTGACAGTGACTACCC 2831
QY 851 CCCACC 856
DB 2832 CCAACC 2837

RESULT 11
US-08-783-774-1
Sequence 1, Application US/08783774
Patent No. 6054130
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
APPLICANT: Jackman, Winthrop
TITLE OF INVENTION: NON-SPICING VARIANTS OF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/783,774
FILING DATE: 15-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Cornuzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7682-037
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5931 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1014...3734
OTHER INFORMATION:

US-08-783-774-1
Query Match 3.9%; Score 54.8; DB 3; Length 5931;
Best Local Similarity 48.7%; Pred. No. 2.4e-06;
Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 551 AGCATATCCAAATTGATCTATGAGCCCAACCAAGATATACAGTCCACCAAGGGA 610
DB 2532 ACCAGCTCACTCAACGATGACTACCCCAACCCCAATGCAACCAAGCCCAACCA 2591
QY 611 ATATGAACTCCACCAAGAGATATGAGCCCAACCAAGGAGATATGAGCCCACTAT 670
DB 2532 GTGACTACCCCAACCCCAATGCAACGAGCCCAACCAAGTACTACCCCAACCA 2651
QY 671 GGGATATGAGAGCCCGCTGTGGATATGAGTCCCACTGGGGATATGAGTCCCA 730
DB 2652 AATGCCACACGCCCCCACTTGGGAAAAACAAGTCTACTGACAGTGACTACCCCAAC 2711
QY 731 TGGGGATATGAGTCCCACTGGGGATATGAGGCCCCCACTGGGGATATGAGTCC 790
DB 2712 CCAATGCCACACGCCCCCACTTGGGAAAAACAAGTCTACTGACAGTGACTACCCCA 2771
QY 791 ACTGGGGATATGAGTCCCACTGGGGATATGAGGCCCCCACTGGAGATATGAGAC 850
DB 2772 ACCCAATGCCACACGCCCCCACTTGGGAAAAACAAGTCTACTGACAGTGACTACCC 2831
QY 851 CCCACC 856
DB 2832 CCAACC 2837

RESULT 12
US-09-556-706B-1
Sequence 1, Application US/09556706B
Patent No. 6458364
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
APPLICANT: Jackman, Winthrop
TITLE OF INVENTION: NON SPICING VARIANTS OF GP350/220
FILE REFERENCE: 7682-050-899
CURRENT APPLICATION NUMBER: US/09/556,706B
CURRENT FILING DATE: 2000-04-24
PRIOR APPLICATION NUMBER: 08/783,774
PRIOR FILING DATE: 1997-01-15
PRIOR APPLICATION NUMBER: 08/229,291
PRIOR FILING DATE: 1994-04-18
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 5931
TYPE: DNA
ORGANISM: Virus
FEATURE:
OTHER INFORMATION: SP350/220
US-09-556-706B-1

Query Match 3.9%; Score 54.8; DB 4; Length 5931;
Best Local Similarity 48.7%; Pred. No. 2.4e-06;
Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 551 AGCATATCCAAATTGATCTATGAGCCCAACCAAGATATACAGTCCACCAAGGGA 610
DB 2532 ACCAGCTCACTCAACGATGACTACCCCAACCCCAATGCAACCAAGCCCAACCA 2591
QY 611 ATATGAACTCCACCAAGAGATATGAGCCCAACCAAGGAGATATGAGCCCACTAT 670
DB 2532 GTGACTACCCCAACCCCAATGCAACGAGCCCAACCAAGTACTACCCCAACCA 2651
QY 671 GGGATATGAGAGCCCGCTGTGGATATGAGTCCCACTGGGGATATGAGTCCCA 730
DB 2652 AATGCCACACGCCCCCACTTGGGAAAAACAAGTCTACTGACAGTGACTACCCCAAC 2711
QY 731 TGGGGATATGAGTCCCACTGGGGATATGAGGCCCCCACTGGGGATATGAGTCC 790

Db 2712 CCAATGCGACGAGCCCACTTGGGAAAAAAGAGCCCACTGAGAGTGAATACCCA 2771
QY 791 ACCCTGGGGGATATGATGATGAGAGCCCACTGAGAGTGAATGAGC 850
Db 2772 ACCCGAATGCGACGAGCCCACTTGGGAAAAAAGAGCCCACTGAGAGTGAATAC 2831
QY 851 CCCAGC 856
Db 2832 CCAAGC 2837

RESULT 13

US-08-182-175A-56
; Sequence 56: Application US/08182175A
; Patent No. 5559223
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/182,175A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Akamehly Floyd
; REGISTRATION NUMBER: 33,692
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; STRAIN: E. coli
; CELL TYPE: DH5 alpha
; IMMEDIATE SOURCE:
; CLONE: 2-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..235
; OTHER INFORMATION: /function= "synthetic storage protein"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /gene= "sap"
; OTHER INFORMATION: /standard_name= "7.7.7.7.7.8.9.8.9.5"
US-08-182-175A-56

Query Match 3.8%; Score 53.4; DB 1; Length 243;
Best Local Similarity 53.0%; Pred. No. 9,3e-07;
Matches 114; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 634 ATGAGCCCAACGAGGGGATATGAGCCCACTATGGAATGAGAGCCCGCTGTGG 693
Db 2 ATGAGAGAGAGAGCTGAAGGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61
QY 694 GATATGAGATCCCACTGGGGGATATGAGATCCCACTGGGGGATATGAGATCCCACTG 753
Db 62 GCATGAGAGAGAGAGCTGAAGGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGCTG 121
QY 754 GGGGATATGAGAGCCCACTGGGGGATATGAGATCCCACTGGGGGATATGAGATCCCACTG 813
Db 122 AAGCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 181
QY 814 CTGGGGGATATGAGAGCCCACTGAGAGATGGA 848
Db 182 CTCAAAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 216

RESULT 14

US-08-474-633A-74
; Sequence 74: Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
; APPLICANT: COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND
; TITLE OF INVENTION: METHODS FOR INCREASING
; TITLE OF INVENTION: INCREASING THE LYSINE
; TITLE OF INVENTION: AND THREONINE CONTENT
; TITLE OF INVENTION: OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,633A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGEL
; REGISTRATION NUMBER: 30,684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; STRAIN: E. coli
; CELL TYPE: DH5 alpha
; IMMEDIATE SOURCE:
; CLONE: 2-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..235
; OTHER INFORMATION: /function= "synthetic
; OTHER INFORMATION: storage protein"

OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "asap"
OTHER INFORMATION: /strand name=
OTHER INFORMATION: "7.7.7.7.7.8.9.8.9.5"
US-08-474-633A-74

Query Match 3.8%; Score 53.4; DB 1; Length 243;
Best Local Similarity 53.0%; Pred. No. 9.3e-07;
Matches 114; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 634 ATGAGCCCAACAGGGGATATGAGCCCACTTATGGAATGAGCCCGCTGTGG 693
DB 2 ATGAGAGAGAGCTGAAGGCGATGAGAGAACTTAAGCCATGAGAGAGAACTGAAG 61
QY 694 GATATGAGTCCCACTTGGGGATATGAGTCCCACTTGGGGATATGAGTCCCACTG 753
DB 62 GCATGAGAGAGAAAGCTGAAGGCGATGAGAGAAAGCTGAAGGCGATGAGAGAAAGCTG 121
QY 754 GGGATATGAGAGCCCACTTGGGGATATGAGTCCCACTTGGGGATATGAGTCCCACTG 813
DB 122 AAGGCGATGAGAGAAAGCTTAAGAAAGATGAGAAAGCTGAATGATGAGAGAGAA 181
QY 814 CTGGGGATATGAGAGCCCACTTGCAGATATGGA 848
DB 182 CTCAAAAAGATGAGAGAAAGCTTAATGATGATGA 216

RESULT 15

US-08-823-771-74
Sequence 74, Application US/08823771
Patent No. 6459019

GENERAL INFORMATION:

APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY

TITLE OF INVENTION:

CHIMERIC GENES AND METHODS FOR INCREASING THE LYSINE AND THREONINE CONTENT

NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESS:

ADDRESSER: E. I. DU PONT DE NEMOURS AND COMPANY

STREET: 1007 MARKET STREET

CITY: WILMINGTON

STATE: DELAWARE

COUNTRY: U.S.A.

ZIP: 19898

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MICROSOFT WORD VERSION 2.0C

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/823,771

FILING DATE: 24-Mar-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/474,633

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BARBARA C. SIEGEL

REGISTRATION NUMBER: 30,684

REFERENCE/DOCKET NUMBER: BB-1037-C

TELEPHONE: 302-992-4931

TELEFAX: 302-773-0164

TELEX: 835420

INFORMATION FOR SEQ ID NO: 74:

SEQUENCE CHARACTERISTICS:

LENGTH: 243 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DHS alpha
IMMEDIATE SOURCE:
CLONE: 2-9

FRATURE:
NAME/KEY: CDS
LOCATION: 2..235
OTHER INFORMATION: /function= "synthetic"

storage protein
/product= "protein"
/gene= "asap"
/strand name=

SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-08-823-771-74

Query Match 3.8%; Score 53.4; DB 4; Length 243;
Best Local Similarity 53.0%; Pred. No. 9.3e-07;
Matches 114; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 634 ATGAGCCCAACAGGGGATATGAGCCCACTTATGGAATGAGCCCGCTGTGG 693
DB 2 ATGAGAGAGAGCTGAAGGCGATGAGAGAACTTAAGCCATGAGAGAGAACTGAAG 61
QY 694 GATATGAGTCCCACTTGGGGATATGAGTCCCACTTGGGGATATGAGTCCCACTG 753
DB 62 GCATGAGAGAGAAAGCTGAAGGCGATGAGAGAAAGCTGAAGGCGATGAGAGAAAGCTG 121
QY 754 GGGATATGAGAGCCCACTTGGGGATATGAGTCCCACTTGGGGATATGAGTCCCACTG 813
DB 122 AAGGCGATGAGAGAAAGCTTAAGAAAGATGAGAAAGCTGAATGATGAGAGAGAA 181
QY 814 CTGGGGATATGAGAGCCCACTTGCAGATATGGA 848
DB 182 CTCAAAAAGATGAGAGAAAGCTTAATGATGATGA 216

Search completed: December 16, 2003, 17:36:27
Job time: 103.547 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: December 16, 2003, 10:43:10 (Search time 465.02 Seconds)

(without alignments)
10099.015 Million cell updates/sec

Title: US-09-864-291-4

Perfect score: 1413
1 cgccacgagggcgccgacgga.....gcataaaaaaaaaaaaaa 1413

Sequence:

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2201672 seqs, 166179599 residues

4403344

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing:

Database:

Published Applications NA:
1: /cgm2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgm2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgm2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgm2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgm2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgm2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgm2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgm2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgm2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgm2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgm2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgm2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgm2_6/ptodata/2/pubpna/US09_NEW_PUBCOMB.seq.*
14: /cgm2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgm2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgm2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
17: /cgm2_6/ptodata/2/pubpna/US60_NEW_PUBCOMB.seq.*
18: /cgm2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	262.6	18.6	467	11	US-09-764-891-1230
2	241.2	17.1	436	9	US-09-864-761-13561
3	220	15.6	7099	11	US-09-764-891-7570
4	195.4	13.8	321	11	US-09-764-891-1356
5	122	8.6	432	11	US-09-918-995-13536
6	116.4	8.2	409	10	US-09-960-352-9984
7	92.4	6.5	291	11	US-09-764-891-7571
8	86	6.1	411	9	US-09-864-761-15029
9	78.6	5.6	531	13	US-10-029-386-24981
10	77.8	5.5	593	13	US-10-027-632-290832
11	77.8	5.3	593	14	US-10-027-632-290832
12	75.2	5.3	420	10	US-09-974-300-3626
13	69.4	4.9	866	13	US-10-029-386-22839
14	68.8	4.9	591	9	US-09-864-761-19727
15	68.8	4.9	1959	9	US-09-864-761-2946

16	68.6	4.9	2016	10	US-09-938-842A-2004	Sequence 2004, Ap
17	64.4	4.6	5769	11	US-09-764-891-7572	Sequence 7572, Ap
18	63.6	4.5	1040	13	US-10-029-386-26033	Sequence 26033, A
19	61.6	4.4	541	13	US-10-029-386-11265	Sequence 11265, A
20	59.8	4.2	3489	13	US-10-294-804-1	Sequence 1, Appl1
21	59.4	4.2	541	13	US-10-029-386-12333	Sequence 12333, A
22	58.8	4.2	417	9	US-09-864-761-27093	Sequence 27093, A
23	58.8	4.2	3048	13	US-10-027-632-113439	Sequence 113439, A
24	58.8	4.2	3048	14	US-10-027-632-113439	Sequence 113439, A
25	57.2	4.0	511	13	US-10-029-386-24981	Sequence 24981, A
26	55.6	3.9	628	13	US-10-029-386-22859	Sequence 22859, A
27	54.8	3.9	2108	10	US-09-962-832-225	Sequence 225, App
28	54.6	3.9	417	9	US-09-864-761-27093	Sequence 27093, A
29	54.6	3.9	422	10	US-09-738-973-337	Sequence 337, App
30	54.6	3.9	422	10	US-09-854-133-337	Sequence 337, App
31	54.6	3.9	422	15	US-10-144-649A-337	Sequence 337, App
32	54.6	3.9	3048	13	US-10-027-632-113439	Sequence 113439, A
33	54.6	3.9	3048	14	US-10-027-632-113439	Sequence 113439, A
34	53.8	3.8	541	13	US-10-029-386-11265	Sequence 11265, A
35	53.4	3.8	243	15	US-10-023-066A-74	Sequence 74, Appl1
36	53.2	3.8	3773	9	US-09-925-302-47	Sequence 47, Appl1
37	53.2	3.8	6354	15	US-10-084-817-158	Sequence 158, App
38	52.6	3.7	354	9	US-09-864-761-18194	Sequence 18194, A
39	52.6	3.7	421	13	US-10-029-386-14222	Sequence 14222, A
40	52.6	3.7	597	13	US-10-029-386-517	Sequence 517, App
41	52.6	3.7	2030	10	US-09-880-107-2416	Sequence 2416, App
42	52.6	3.7	8923	10	US-09-764-847-1041	Sequence 1041, Ap
43	52.6	3.7	8923	15	US-10-092-154-1041	Sequence 1041, Ap
44	52.4	3.7	2746	13	US-10-371-725-1	Sequence 1, Appl1
45	52.2	3.7	689	13	US-10-027-632-148124	Sequence 148124, A

ALIGNMENTS

RESULT 1
US-09-764-891-1230
US-09-764-891-1230, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PAM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1230
LENGTH: 467
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: SITE
LOCATION: (421)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1230

Query Match	18.6%, Score 262.6, DB 11, Length 467, Best Local Similarity 76.4%, Pred. No. 4.6e-74, Matches 365, Conservative 0, Mismatches 100, Indels 13, Gaps 3,
DB	386 CAGGAAAGAGAGTCCATCGAATTCGCCCACTGATGTAAAGCTGCTGCTGCTGC 445
QY	2 CAGGAAAGAGAGTCCATCGAATTCGCCCACTGATGTAAAGCTGCTGCTGCTGC 61
DB	446 CAGGAAATTCACCTGGAAGTGAATTAATTCGTTGGAACCTTCAGACTGTCATAT 505
QY	62 CCGAGAAATTCACCTGGAAGTGAATTAATTCGTTGGAACCTTCAGACTGTCATAT 121
DB	506 TACTGTCCAGAGGCTGAGTGTCTCTCCACAGACACTTGTCCAGACTATTCATAT 565
QY	122 TACTGTCCAGAGGCTGAGTGTCTCTCCACAGACACTTGTCCAGACTATTCATAT 169

```

Qy 566 GATCTATGAGCCCCCAGCAGATATACAGTCCAAACGAGGGAATATGAACTCCACC 625
Db 170 TGTCTATGAGCCCCCAGCAGATATGAGAGCCCACTCCCGGATACGAGAGCCCAACC 229
Qy 626 AGAAGATATGAGAGCCCAACGAGGGAATATGAGAGCCCACTATGAGATATGAGAGCC 685
Db 230 TGCAGATATGAGAGCCCAACGAGGGAATATGAGAGCCCACTATGAGATATGAGAGCC 289
Qy 686 GCGTGTGAGATATGAGAGTCCCACTGAGGGAATATGAGATCCCACTGAGGGAATATGAG 745
Db 290 ACTGTGCAATATGAGAGCCCACTTGTGATACGAGAGCCCACTGAGGGAATATGAGAGC 349
Qy 746 CCCACCTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGATCCCACTGAGGGAATATG 804
Db 350 CCCACCTGAGGGAATATGAGAGCCCACTTGTGATACGAGAGCCCACTGAGGGAATATG 409
Qy 805 GAGCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGATCCCACTGAGGGAAT 862
Db 410 GAGCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGATCCCACTGAGGGAAT 467

```

RESULT 2

```

US-09-864-761-31561
; Sequence 31561, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117

```

```

; SOFTWARE: Annonex Sequence Listing Engine vers. 1.1
; SEQ ID NO 31561
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Homo sapiens
; PRIMER:
; OTHER INFORMATION: MAP TO 299716.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
; OTHER INFORMATION: EST HUMAN HIT: AM169980.1, EVALU8 8.00e-11
; OTHER INFORMATION: SWISSPROT HIT: P18616, EVALU8 3.00e-13
US-09-864-761-31561

```

```

Query Match 17.1%; Score 241.2; DB 9; Length 436;
Best Local Similarity 73.2%; Pred. No. 3.8e-67;
Matches 328; Conservative 0; Mismatches 108; Indels 12; Gaps 1;

```

```

Qy 562 TTGTGATCTATGAGAGCCCAACGAGGGAATATGAGAGCCCACTATGAGATATGAGAG 621
Db 1 TTATGTCTATGAGAGCCCACTGAGGGAATATGAGAGCCCACTCCGATACGAGAGCC 60
Qy 622 CACGAGAGATATGAGAGCCCAACGAGGGAATATGAGAGCCCACTATGAGATATGAGAG 681
Db 61 CACTGAGAGATATGAGAGCCCAACGAGGGAATATGAGAGCCCACTGAGATATGAGAGAG 120
Qy 682 CCCCCCTGTGAGATATGAGATCCCACTGAGGGAATATGAGATCCCACTGAGGGAATATG 741
Db 121 CCTCACCTGTGAGATATGAGAGCCCACTCTTGTGATACGAGAGCCCACTGAGATATG 180
Qy 742 GAGTCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGATCCCACTGAGGGA 801
Db 181 GAGCCCACTGAGGGAATATGAGAGCCCACTCTTGTGATATGAGAGCCCACTCTGAGAT 240
Qy 802 ATGTGTCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTGAGG 861
Db 241 ATGAGAGCCCACTCTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTG 300
Qy 862 GAAATGAGAGCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTG 921
Db 301 GATACAGAGCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTG 356
Qy 922 ACAGATCTATGAGAGCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGAG 981
Db 357 -----CAGGCTCTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTG 408
Qy 982 ATTACCACTCTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTG 1009
Db 409 ATTCTTAACCTCTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTG 436

```

RESULT 3

```

US-09-764-891-7570
; Sequence 7570, Application US/09764891
; Publication No. US2003007808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7570
; LENGTH: 7099
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7570

```

```

Query Match 15.6%; Score 220; DB 11; Length 7099;
Best Local Similarity 75.3%; Pred. No. 1.4e-59;

```

	Matches	274; Conservative	0; Mismatches	90; Indels	0; Gaps	0;
OY	554	ATATTCAAATTGTGATCTTATGGAACCCCAACACAGAGATATACGTCCACACAGGGGAAATA				613
Db	6718	ATTCCAGATATATGCTATATGAGACCCCACTGACAGATATATGAGACCCCACTCCCGGATTA				6777
OY	614	TGGAATCTCAACAGAGGATATGAGACCCCAACAGGGGGATATGAGACCCCACTTATGGG				673
Db	6778	CGAGACCCCACTGACAGGATATGAGACCCCAACCTGAGAAATGAAAGGCCCGGCTGTGGG				6837
OY	674	ATATGAGACCCCGGCTGTGGGATATGAGAGTCCCACTCTGGGGGATATGAGATGCCACTGG				733
Db	6838	ATACAGAGCTTCACCTGTGGCATATGAGACCCCACTCTTGGAATCGAGGCCCACTGCG				6897
OY	734	GGGATATGAGATGCCCACTGTGGGGATATGAGACCCCACTTGGGGGATATGAGATGCCCAAC				793
Db	6898	AGGATATGAGACCCCACTTACAGATATGAGACCCCACTCTTGGATATGAAACCCCAAC				6957
OY	794	TGGGGGATATGATGCCCACTGGGGGATATGAGAGCCCACTGTGAGATATGAGAGGCC				853
Db	6958	TCTCGGATATGAGGCCCACTCTGGGATATGAGACCCCACTGTGAGAAATGAAAGGCC				7017
OY	854	ACGAGCTGGAATGAAAGCCTTACCCCTGCATATGAGATCTCATCTGCTGGAATAACGC				913
Db	7018	GCTCTGGGATACAGAGCTTCACCTGTGAGATCAGAGACAGGCTTCAGGAATCTACAGC				7077
OY	914	TGCC				917
Db	7078	AGCC				7081

```

RESULT 4
US-09-764-891-1356
; Sequence 1356, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1356
; LENGTH: 321
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: SITE
; LOCATION: (316)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (321)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1356

Query Match      13.8%; Score 195.4; DB 11; Length 321;
Blast Local Similarity 77.4%; Freq. No. 2.2e-52;
Matches 246; Conservative 2; Mismatches 69; Indels 1; Gaps 1,

Qy      64  GCCGCGAGGGGACCTCATCCCTCTGGCGAAGTCTGTGAAGCAGGTGAGATGCG 123
Db      1  GCCGCCGCGAGACCTCATCCCTTAACGTGAAAGTCTTTAAGCGAGTCCGAATGTGG 60

Qy      124 ACCTTCGCTTCTTAAGAAACGATGGAATCTATCTCTTAATGCAAAAGAAAGGA 183
Db      61  AGCTCTCTTCCACAGGATCAGAAAGCTCAAAATGTCTTTAGTGTGAAGACAGGAA 120

Qy      184 CGTTGTTTCTCACTTCAATACCGGGGAGTCTTGTGACCTTCACTTAAGTCAATGACCCA 243
Db      121  CATTGTTTCTCACTTCATACCGGGTGAATTTCTAATCACTTCAATGCTCCATGATGATCCA 180

```

Qy	244	TCGCTCTTTAAGAGCGGCTTGCGCTGATGATGACGACATTTGAAACAACATTT	303
Db	181	TGTGTCTTTAAGAGCCATTGTATGTGAATGACAACTCACTGTGGAACAACTAT	240
Qy	304	TTGCCCCAATCATTTAAGGACATTCAGGAGCTCCAGGTGTGTGGCTGGG- AAGGA	362
Db	241	TTTGCTGCAACTTCATTAAAGGAACTATTCAAGCACTTCATATGAGGCTGGGAAGGA	300
Qy	363	CAAGCTGTTTTAAGTTA	380
Db	301	CAAGCTACTTTAAANTA	318

```

RESULT 5
US-09-918-995-33536
Sequence 33536, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 33536
LENGTH: 432
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(432)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-33536

```

Query Match	Similarity	8.6%	Score 122	DB 11	Length 432
Best Local	Similarity	59.4%	Pred. No. 1.5e-26		
Matches 247	Conservative	0	Mismatches 160	Indels 9	Gaps 2
Qy	36	ATGCACTGAAACAGAGCCACACGAGAGCGCGTGTGTGGGCGCTCATCCCTCTGGCGAA	95		
Db	18	ATGGCGCTCAACAGAAATCACTGGAGGGGGGCGAGATGTGTCAATTAACAC---CGAG	74		
Qy	96	AGTGTCTTGAAGCAGTGTGAGATGTGACCTCTGTCTTCTTACAGAAACAGTGAATCC	155		
Db	75	AGCATCCTAATGGCTTATGATCAGTGTGAATCTACATTCAATGACATGAAGACGTGCA	134		
Qy	156	TATCTTTTATATGSCACAAAGAAAGAACGTGTGTTCTACCTTCAATACCGGATGGCTTC	215		
Db	135	GAACCTTTCAAAGGACCAAGAAAGGACGTGTCTACCTTAACCCCTTACCGGATCATCTTT	194		
Qy	216	GTCATCTTCACTTAATGTAATGACCCCAATGCTTTCTTTATATATGATGCGGTTGGCGTAATG	275		
Db	195	CTGTCT-----CAAGGCGCAAGAGATGCATGACGTCTTCAATGATGCACTTTATATCTCAATG	248		
Qy	276	AGTACCTGCACCAATTGAACCAACAAATTTTGGCCCCCAACTACATTAAGAAACCAATTGAG	335		
Db	249	AAAGACTGTGATATCAAGCAAGCCCGTATTTGTGTCAAACTACATCAAGAGAAACAGTGAAG	308		
Qy	336	GCAAGCTCAAGGTGTGCTGTGGAGAGAACAGAGCTGTTTTTAAGTATCTCTTCAAGAAAGAA	395		
Db	309	GGCGAAGGGGAGGTGTGCTGTGGAGAGGCTCTGTCTTCTAACAAGTTGACCTTTCAAGGCAAGG	368		
Qy	396	GGTGCATTCGAATTTGGCCCACTGATGTGTAAAGAGTGCCTTGTCTGTCCAGAGG	451		
Db	369	GGGCGCAATGTAGTTGTGACACAGGATGTCTCAAGGTGTGCATCTCAAGGCTTCAGAGG	424		

RESULT 6
US-09-960-352-6984

Sequence 6984, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Mengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE OF INVENTION: MOSCIE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 6984
LENGTH: 409
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 30-LIB3058-024-Q1-K1-H9
US-09-960-352-6984

Query Match 8.2%; Score 116.4; DB 10; Length 409;
Best Local Similarity 59.2%; Pred. No. 9.6e-27;
Matches 239; Conservative 0; Mismatches 156; Indels 9; Gaps 2;

QY 26 CCTGGGCGAGATGAGTGAACCAAGGACCAACCGGAGCCGTCGCGGCGCCCTCATGCC 85
DB 13 CCGAGGAGCTATGGCTCTCATGAAGACCACTCGAGGCGCGGAGTATGCTCAACAA 72
QY 86 CTCTGGCGAAGAGTCTTGAAGCAGTGTGAGATGTGACCTCTGCTCTCTACAGAAAC 145
DB 73 CAC--CGAGAGATCTTATGTCTTATGACCATGATGACCTTACCTTACAGTACAG 129
QY 146 AGTGAATCTTATCTTATATGCAAAAGAAAGACGTTGTTTCTACATTCATACCG 205
DB 130 GAATGCGCAAGAGCCCTTCAAGGAGCAAGAAAGGACCGTACCTTACCCGTAACG 189
QY 206 GGTGGTCTTGTGACTTACACTTATGATGACATGACCCGATGCTTTTATATGAGCCGT 265
DB 190 GGTGATCTTCTGTC-----CAAGCGAGGAGTGCATGCACTCTTCATGATGCCCCTT 243
QY 266 TGGCGATGAGTGAAGTGCACATTCGAAACCAATTTTGGCCCACTACATTAAGG 325
DB 244 CTATCTGATGAGAGCTGTGATGATCAAGACGCTGTGTTGGGCAAACTACATCAAG 303
QY 326 AACCATTCAGGAGCTTCAAGTGTGCTGCTGGAGAGCAAGCTGTTTAACTTATCTT 385
DB 304 GACAGTGAAGCGCGAAGAGAGAGAGTGTGCTGGAGAGCTGCACTACAGTGAACCTT 363
QY 386 CAGGAAAGAGTGCATGCAATTTGCGCCCACTGATGTAAAG 429
DB 364 TATGTTGCGGGGCGCATGAAATATGACAGCGGATGTACAG 407

RESULT 7
US-09-764-891-7571
Sequence 7571, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: P0006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7571
LENGTH: 291
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-891-7571

Query Match 6.5%; Score 92.4; DB 11; Length 291;

Best Local Similarity 76.0%; Pred. No. 4.6e-19;
Matches 114; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 206 GGTGCTTCGTGACTTCACTTATGATGACCCCAATGCTTTCTTATGATGCCGT 265
DB 9 GGTGATTTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 68
QY 266 TGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 325
DB 69 TATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 128
QY 326 AACCATTCAGGAGCTTCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 355
DB 129 AACATTCAGGAGCTTCACTTATGATGATGATGATGATGATGATGATGATGATGATGAT 158

RESULT 8
US-09-864-761-15029
Sequence 15029, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wenheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 15029
LENGTH: 471
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:

OTHER INFORMATION: MAP TO Z99716.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
US-09-864-761-15029

Query Match 6.1%; Score 86; DB 9; Length 471;
Best Local Similarity 77.6%; Pred. No. 7.3e-17;

Matches 104; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 662 CCCACCTATGAGATGAGAGCCCGCTGTGAGATATGAGATCCACCTGGGGATATG 721
DB 338 CCCAGTATATGCTATGAGAGCCCGCTGTGAGATATGAGATCCACCTGGGGATATG 397
QY 722 AGTCCACCTGGGGATATGAGATCCCGCTGTGAGATATGAGATCCCGCTGGGGAT 781
DB 398 AGCCCGCTGTGAGATATGAGAGCCCGCTGTGAGATATGAGATCCCGCTGGGGAT 457
QY 782 TGGAGTCCCGCTGT 795
DB 458 CAGAGCTTCACCTG 471

RESULT 9

US-10-029-386-24981
Sequence 24981, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE OF INVENTION: EXPRESSION ANALYSIS TWO
FILE REFERENCE: ABOVICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annonex Sequence Listing Engine vers. 1.1
SEQ ID NO 24981
LENGTH: 511
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO Z97205.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
US-10-029-386-24981

Query Match 5.6%; Score 78.6; DB 13; Length 511;
Best Local Similarity 49.9%; Pred. No. 1.9e-14;

Matches 198; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

QY 520 CTGAGTGTGCTCTCAACAGACCTTGTCCAGCATATCCATTTGATATGAGCC 579
DB 106 CTGAGAGGCCCCCTCTCTCACTGAGAGTCCCCCTCTCTCACTGAGAGTCCCCCTCTC 165
QY 550 CACCAACAGATATACAGTCCAAACAGAGGAAATATGAACTCCACCAAGAGATATGAG 639
DB 166 CACTGAGAGCTCCCTCTCTCACTGAGAGTCCCCCTCTCTCACTGAGAGTCCCCCTC 225
QY 640 CCGAACAGAGGAGATATGAGAGCCCGCTGTGAGATATGAGAGCCCGCTGTGAGATATG 699
DB 226 CTGAGCTGAGAGTCTCTCTCTCACTGAGAGTCCCCCTCTCTCACTGAGAGTCTCT 285
QY 700 GAGTCCACCTGGGGATATGAGATCCCGCTGTGAGATATGAGATCCCGCTGGGGAT 759
DB 286 CTCTCTCACTGAGAGTCTCTCTCTCACTGAGAGTCTCTCTCTCACTGAGAGTCTCT 345
QY 760 ATGAGAGCCCGCTGTGAGATATGAGATCCCGCTGTGAGATATGAGATCCCGCTGGGG 819
DB 346 CCGTCTCTCACTGAGAGTCTCTCTCTCACTGAGAGTCTCTCTCTCTCACTGAGAG 405
QY 820 GATATGAGAGCCCGCTGTGAGATATGAGAGCCCGCTGTGAGATATGAGAGCCCGCT 879

DB 406 GCTCCCTCTCTCACTGAGAGTCTCTCTCTCACTGAGAGTCCCCCTCTCACTG 465
QY 880 CTGATATGAGAGTCTCTCTCTCTGAAATATGAGTCTC 916
DB 466 CAGGCTCCCTCTCTCACTGAGAGTCTCTCTCTCTC 502

RESULT 10

US-10-027-632-290832
Sequence 290832, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 290832
LENGTH: 593
TYPE: DNA
ORGANISM: Human
US-10-027-632-290832

Query Match 5.5%; Score 77.8; DB 13; Length 593;
Best Local Similarity 54.9%; Pred. No. 3.8e-14;
Matches 151; Conservative 1; Mismatches 122; Indels 0; Gaps 0;

QY 177 AAAGAACTGTTTCTCACTTACCGGGTGTCTGTGACTTCACTTATGATCAAT 236
DB 45 AAAGAACTGTTTCTCACTTACCGGGTGTCTGTGACTTCACTTATGATCAAT 236
QY 237 GACCCAGCTTCTTATATGATGCGGCTTGTGATGATGATGATGATGATGATGATG 296
DB 105 GATGCAAGAGTCTCTGTATATGATGATGATGATGATGATGATGATGATGATGATG 164
QY 297 CCAATTTTGGCCCCCACTTATTAAGAAACATTCAGGAGCTCCAGGTGGTGGCTGG 356
DB 165 CCGTGTATGACAACTGATCAAGAGAACTGATTAACAGAGAGAGAGAGAGAGAG 224
QY 357 GAGGACAACTGTTTATGATATCTTCAAGAAAGAGGTGTCATGAAATTTGCCCA 416
DB 225 GAGGCTGCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 284
QY 417 CTGATGTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 451
DB 285 CAGGTCTCCAGGTGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 319

RESULT 11

US-10-027-632-290832
Sequence 290832, Application US/10027632
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome

```

FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 290832
LENGTH: 593
TYPE: DNA
ORGANISM: Human
US-10-027-632-290832

```

```

Query Match 5.5%; Score 77.8; DB 14; Length 593;
Best Local Similarity 54.9%; Pred. No. 3.8e-14;
Matches 151; Conservative 1; Mismatches 123; Indels 0; Gaps 0;

```

```

QY 177 AAGGAAAGCTTTTCTCACTTCAATCCGGGTGCTCTTCTGACTTCACTTCAAT 236
    |||||
DB 45 AAGGAGCCAAAGGACACATCTTACCCCTTCACTTCTTCTGCTCAAGGAG 104
    |||||
QY 237 GACCCATGCTTTTATGATGCGTTGGCTGATGATGATGATGATGATGATGATGAT 296
    |||||
DB 105 GATGCCACGAGTCTCTGTAATGATCTTATTTATTTGAAAGCTTATGATCAAGCAG 164
    |||||
QY 297 CCAATTTTCCCACTAATTAAGAACCAATTCAGGAGCTCCAGGTGCTGCTG 356
    |||||
DB 165 CCGTGTGTTACCAATCTCATTAAGGAACGTAACCCAGAGCGGCTGCTG 224
    |||||
QY 357 GAAGACAGCTGTTTAAATGATCTTCAAGGAAGAGGTGCTCAATTTGCCCA 416
    |||||
DB 225 GAAGGCTCTCTCTGCGAGTCCGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAG 284
    |||||
QY 417 CTGATGCTAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 451
    |||||
DB 285 CAGTCTCCAGGATGATCTCAAGGCTGCAAGGAGGAGGAGGAGGAGGAGGAGGAG 319
    |||||

```

```

RESULT 12
US-09-974-300-3626
Sequence 3626, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3626
LENGTH: 420
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-3626

```

```

Query Match 5.3%; Score 75.2; DB 10; Length 420;
Best Local Similarity 49.7%; Pred. No. 2.1e-13;
Matches 191; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

```

```

QY 557 TCCAAATGATCTATGACCCCAACCAAGATATCAATCAAGGAGGATATG 616
    |||||
DB 8 TCGAAGCGGCTATATGAGTCAAGCGGCTGATATGAGTCAAGCGGCTGATATG 67
    |||||
QY 617 AACTCCACCAAGAGATATGAGCCCAACAGGAGGATATGAGCCCACTATGAGATA 676
    |||||
DB 68 CAGTCAAGCGGCTATATGAGTCAAGCGGCTGATATGAGTCAAGCGGCTGATATG 127
    |||||
QY 677 TGAAGCCCGCTGATGATATGAGTCCCACTGAGGAGATATGAGTCCCACTGAGG 736
    |||||
DB 128 TGAAGTCAAGCGGCTGATATGAGTCAAGCGGCTGATATGAGTCAAGCGGCTG 187
    |||||
QY 737 ATATGAGTCCCACTGAGGAGATATGAGCCCACTGAGGAGATATGAGTCCCACTG 796
    |||||
DB 188 ATATGAGTCCCACTGAGGAGATATGAGCCCACTGAGGAGATATGAGTCCCACTG 247
    |||||
QY 797 GGGATATGATGAGCCCACTGAGGAGATATGAGCCCACTGAGGAGATATGAGCCCA 856
    |||||
DB 248 CTGATATGATGAGTCAAGCGGCTGATATGAGTCAAGCGGCTGATATGAGTCAAG 307
    |||||
QY 857 AGCTGGAATGAGCCCTTACCCCTGATATGAGTCAATCTGCTGAGAAATACAGTGC 916
    |||||
DB 308 CCGCTGATATGAGTCAAGCGGCTGATATGAGTCAATCTGAGGAGATATGAGTCAAG 367
    |||||
QY 917 CTCTCAAGATATGAGTCAAGTCA 940
    |||||
DB 368 AACCGATATGAGTCAAGTCAAGTCA 391
    |||||

```

```

RESULT 13
US-10-029-386-22839/C
Sequence 22839, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AEOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 22839
LENGTH: 866
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
OTHER INFORMATION: MAP TO AC005630.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
OTHER INFORMATION: SWISSPROT HIT: 008379, EVALUR 4.00e-10
OTHER INFORMATION: NT HIT: g116161055, EVALUR 4.00e-95
OTHER INFORMATION: EST_HUMAN HIT: BF828725.1, EVALUR 0.00e+00
US-10-029-386-22839

```

```

Query Match 4.9%; Score 69.4; DB 13; Length 866;
Best Local Similarity 49.9%; Pred. No. 2.5e-11;
Matches 175; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

```

```

QY 571 ATGACCCCAACCAAGATATGAGTCAAGCGGAGATATGAGTCAAGCGGAG 630
    |||||
DB 509 ACCTAGCTCTCTCTGTTCAACAGGCTCTCTCTGTTCAAGTCTCTCTCTG 450
    |||||
QY 631 GATATGAGCCCAACAGGAGATATGAGCCCACTATGAGTATGAGCCCGCTG 690
    |||||

```


Job time : 469.02 secs

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2946
; LENGTH: 1959
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; OTHER INFORMATION: MAP TO AC002041.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 4.3
US-09-864-761-2946
```

```

Query Match 4.9%; Score 68.8; DB 9; Length 1959;
Best Local Similarity 51.6%; Pred.No.6.2e-11;
Matches 157; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
```

```

QY 602 ACCAGGGGATATGGAATCCAGCAGAGATATGAGCCCAACGAGGGGATATGAGC 661
   |||||
DB 479 ACCTGGGGTATGTGAGGTGACGTGGGGAATGTGTGAGGTGACCTGGGGGACGTGTAG 538
   |||||
QY 662 CCACCTATGGAATATGAGCCCGCTGTGGGATATGAGTCCACCTGGGGGATATGG 721
   |||||
DB 539 GTGACCTGTGGGACATGTGAGGTGACCTGGGGGATGTGTGAGGTGACCTGGGGGACGTGT 598
   |||||
QY 722 AGTCCACCTGGGGGATATGAGTCCACCTGGGGGATATGAGGCCCACTGGGGGATA 781
   |||||
DB 599 GAGGTGACCTGGGGATGTGTGAGGTGACCTGTGGGACATGTGTGAGGTGACGTGTGACATG 658
   |||||
QY 782 TGAAGTCCACCTGGGGGATATGTGTCGCCCACTGGGGGATATGAGGCCCACTGTGCAGG 841
   |||||
DB 659 TGTGAGGTGACGTGGGGAATGTGTGAGGTGACCTGGGGGACGTGTGACGTGACGTGGGGA 718
   |||||
QY 842 ATATGAGCCCAACCACTGTGAATGAAGCCCTACCCCTGCATATGAAGCTCCATCTGC 901
   |||||
DB 719 ATGTGTGAGGTGACCTGGGGGACATGTGTGAGGTGACCTGGGGGATGTGTGAGGTGACCTGG 778
   |||||
QY 902 TGA 905
   |||
DB 779 GGA 782
```

Search completed: December 16, 2003, 17:58:42

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 09:21:45 / Search time 2907.96 Seconds

(Without alignments)
11809.735 Million cell updates/sec

Title: US-09-864-291-4

Sequence: 1 cggcagcagggcgcgagga.....gcataaaaaaaaaaaaaa 1413

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estda:*
2: em_estdm:*
3: em_estln:*
4: em_estml:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_png:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	546.8	38.7	1079	12	BM564167 AGENCOURT
2	422.4	29.9	1414	11	AK015863 AK015863
3	346.2	24.5	663	10	BC699398 BC699398
4	345.4	24.4	775	10	BG722815 BG722815

5	341.8	24.2	717	14	BY715414	BY715414
6	211.4	15.0	701	10	BG701881	BG701881 602683237
7	197.8	14.0	360	13	BY098622	BY098622 BY098622
8	183.6	13.0	932	13	BU107273	BU107273 603111954
9	182.2	12.9	814	13	BU0961805	BU0961805 AGENCOURT
10	179.4	12.7	500	13	BU103740	BU103740 603005979
11	179.4	12.7	700	13	BU284977	BU284977 603603379
12	179.4	12.7	723	9	AJ452591	AJ452591
13	177.8	12.6	568	12	BU390687	BU390687
14	177.8	12.6	636	12	BM490729	BM490729
15	167.4	11.8	772	9	AJ454950	AJ454950
16	166.8	11.8	706	13	BU433499	BU433499 603257849
17	155.4	11.0	460	10	BB680329	BB680329 dg02h09.y
18	155	11.0	719	14	CB593075	CB593075 AGENCOURT
19	154	10.9	760	14	CB592279	CB592279 AGENCOURT
20	153.8	10.9	862	13	BU735296	BU735296 AGENCOURT
21	153.8	10.9	868	14	CD255258	CD255258 AGENCOURT
22	153.6	10.9	816	13	BU916890	BU916890 AGENCOURT
23	153.6	10.9	931	13	BU914288	BU914288 AGENCOURT
24	152.2	10.8	613	9	AL849211	AL849211
25	152.2	10.8	647	9	AL849651	AL849651
26	152.2	10.8	655	9	AL898723	AL898723
27	152.2	10.8	661	9	AL972453	AL972453
28	152.2	10.8	663	9	AL881810	AL881810
29	152.2	10.8	663	9	AL896682	AL896682
30	152.2	10.8	726	14	CB591559	CB591559 AGENCOURT
31	152	10.8	509	10	BS15416	BS15416 dae07a12
32	152	10.8	594	12	BM190972	BM190972 da130c07
33	152	10.8	607	12	BM190975	BM190975 da130c10
34	151.6	10.7	460	10	BS13187	BS13187 dae06a10
35	150.6	10.7	611	9	AL846063	AL846063
36	150.4	10.6	892	14	CD301781	CD301781 AGENCOURT
37	143.6	10.2	488	28	AQ997933	AQ997933 RPT-23-3
38	141.2	10.0	635	9	AL889954	AL889954
39	136.8	9.7	622	10	BE468886	BE468886 IPHdK0053
40	135.6	9.6	899	10	BE470466	BE470466 IPHdK0305
41	132.8	9.4	652	14	CA590725	CA590725 hab30c06
42	131.8	9.3	631	13	BU805447	BU805447 haa41e09
43	131.8	9.3	744	13	BU230211	BU230211 603398308
44	131.2	9.3	635	13	BU807307	BU807307 haa15a08
45	130	9.2	437	13	BY249862	BY249862

ALIGNMENTS

RESULT 1
LOCUS BM564167
DEFINITION AGENCOURT_6560149 NIH_MGC_119 Homo sapiens CDNA clone IMAGE:5742129
ACCESSION BM564167
VERSION BM564167.1 GI:18811738
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
REFERENCE NIH-MGC http://mgs.nci.nih.gov/
AUTHORS 1 (bases 1 to 1079)
TITLES Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL Bukaryota; Euteleostomi; Mammalia; Euteleostomi; Euteleostomi;
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaab@remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12759 row: m column: 10
High quality sequence stop: 726.

FEATURES
SOURCE

Location/Qualifiers
1. 1079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5742129"
/cissue_type="medulla"
/lab_host="DH10B"
/clone_11b="NIH-MGC_119"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (inventor). Research Genetics tracking code 013. Note: this is a NIH-MGC Library."

BASE COUNT 267 a 297 c 268 g 247 t
ORIGIN

Query Match 38.7%; Score 546.8; DB 12; Length 1079;
Best Local Similarity 73.7%; Pred. No. 4.2e-113;
Matches 772; Conservative 0; Mismatches 252; Indels 24; Gaps 5;

21 GGGGCGCTGGGCGAGATGAGCAAGCAAGCCAGAGAGCGCTGCGGGCCCTC 80
22 GAGGCCCGAAGCAAGGCGGTGATGAGCCAGCAAGCAAGCGCGCGAGCCCTC 91
81 ATCCCTCTGCGGAAAGTCTTTGAAAGAGTGTGAGATGAGATCTGCTTCTCAAG 140
92 ATCCCTTAACGTAAGAGTCTTTGAAAGCGGTCTCGAATGTGAGATCTCTCTCCACAG 151
141 AAACGAGTGAATCTATCTCTTTATGAGCAAGAAAGAAAGAAAGCTGTTCTCACTCA 200
152 CGATGAGAAAGCTCAATGCTCTTTAGTGTGAGAAAGCAAGAACTGTTCTCACTCA 211
201 TACCGGGGTGCTTCTGATGATTCAGATTAAGTCAATGAGCCGATCTTTCTTATGAG 260
212 TACCGGGGTGATTTTCAATGATTCAGATTAAGTCAATGAGCCGATCTTTCTTATGAG 271
261 CGGTGAGCTGATGAGTGTGAGCAATGAGCAAGCAAGCAAGTGTGAGCAAGTGTGAG 320
272 CCATTGATGATGATGAGCAAGTGTGAGCAAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 331
331 AAAGCAATGATGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 380
332 AAGGCAATGATGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 391
381 TCCCTGAGAAAGAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 440
392 GCTTCAAGAAATGAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 451
441 GCTGCGAAGAAATGATTCAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 500
452 GCTGCGCGAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 511
501 ATAAATGATGATGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 560
512 GTAATTAATGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 560
561 ATTGATGATGATGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 620
562 -TTATGATGATGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 619
621 CCACGAGAGATGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 680
620 CCACCTGAGAGATGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 679
681 GCGCGCGCTGAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 740
680 GCTTCACTGAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 739

QY 741 GAG-TGCCCTGAGGAGTATGAGGCCCACTGAGGAGTATGAGATGCCCACTGAGGAG 799
DB 740 GAGGCCCGACCTCTAGATATGAGGCCCACTCTTGAATATGAGGCCCACTCTGAG 799
QY 800 ATATGAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 859
DB 800 ATATGAGGCCCACTCTGAGTATGAGGCCCACTCTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 859
QY 860 TGAATGAGGCCCACTCTGAGTATGAGGCCCACTCTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 919
DB 860 GAGTATGAGGCCCACTCTGAGTATGAGGCCCACTCTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 912
QY 920 TCAAGATGATGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 979
DB 920 TCAAGATGATGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 979
QY 980 CCATTGACCACTCTGAGTATGAGGCCCACTCTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 1038
DB 970 CCATTGACCACTCTGAGTATGAGGCCCACTCTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAGCAAGTGTGAG 1029

QY 1039 AACTGAAATGACAGTAAAGAAAGAAC 1066
DB 1030 AATTGAAAGTGGGAAATGAGAGACC 1057

RESULT 2
AK015863
LOCUS
DEFINITION
MUS musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930521I23 product:hypothetical protein, full insert sequence.
AK015863.1 GI:12854363
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCES
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCES
AUTHORS

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

3 Shibata, K., Itoh, M., Mizawa, K., Nagasaka, S., Sakaki, N., Carninci, P., Kono, H., Akiyama, T., Nishii, K., Kitanishi, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Mashimo, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaue, S., Ikegami, T., Kaishiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Iwase, M., Ohara, B., Matshiki, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumoto, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Iwase, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., Kling, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,

QY 780 TATGAGTCCACCTGAGGATATGATGCCCACTGGGAGATATGAGCCCACTGCA 839
 DB 775 TATGAGTCCACCTGAGGATATGATGCCCACTGGGAGATATGAGCCCACTGCA 834
 QY 840 GATATGAGCCCACTGAGGATATGAGCCCACTGGGAGATATGAGCCCACTGCA 899
 DB 835 GATATGAGCCCACTGAGGATATGAGCCCACTGGGAGATATGAGCCCACTGCA 894
 QY 900 GCTGAAATACAGCTGCTCTCAAGATCT 929
 DB 895 ATGGATCCGGTCCCACTCCAGATAT 924

RESULT 3
 BG69398 663 bp mRNA linear EST 07-MAY-2001
 LOCUS 602679081F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4811804 5',
 DEFINITION mRNA sequence.

ACCESSION BG69398 GI:13967653
 VERSION BG69398
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Butelostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 663)
 NIH-MGC http://imgc.ncl.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-r@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shihrai
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLNL0703 row: a column: 21

High quality sequence stop: 660.
 Location/Qualifiers

FEATURES
 source 1..663

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:4811804"
 /feature_type="hipocampus"
 /lab_host="DH10B"
 /clone_lib="NIH MGC 95"
 /note="Organ: Brain; Vector: pBluescriptPCr (modified
 pBluescript KS+); Site 1: BamHI; Site 2: SalI; XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTA-3',
 size-selected for average insert size 2.5 kb and
 normalized to R0T 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH-MGC library."
 BASE COUNT 168 a 158 c 169 g 164 t 4 others

Query Match 24.5%; Score 346.2; DB 10; Length 663;
 Best Local Similarity 74.7%; Prid. No. 9.8e-68;
 Matches 490; Conservative 0; Mismatches 151; Indels 15; Gaps 4;

QY 21 GGGGCGCTGGGCGAGATGAGTGAACCAAGCACCAGAGAGCGCTGCGGGCGCTC 80
 DB 22 GAGGCGCGGAGAGATGAGTGAACCAAGCACCAGAGAGCGCTGCGGGCGCTC 81
 QY 81 ATCCCTCTGGCGAAGTGTCTTGAAGCAATGTGAGATGAGATGAGATGAGATGAG 140

DB 82 ATCCCTAAGGAGAAAGTGTCTTGAAGCAATGTGAGATGAGATGAGATGAGATGAG 141
 QY 141 AAACAGTGAATCTTATCTTTATGAGCAAAAGAAAGAACTTTTCTCACTTCA 200
 DB 142 CGATCAAGAGGCTCAAAAGTGTCTTGAAGCAATGTGAGATGAGATGAGATGAGATGAG 200
 QY 201 TACCGGGGCTTGTGATCTTCACTTGAAGCAATGTGAGATGAGATGAGATGAGATGAG 260
 DB 201 TACCGGGGCTTGTGATCTTCACTTGAAGCAATGTGAGATGAGATGAGATGAGATGAG 260
 QY 261 CCGTTGGCTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 320
 DB 261 CCGTTGGCTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 320
 QY 321 AAGGAAATATTCAGAGAGCTCCATATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 380
 DB 321 AAGGAAATATTCAGAGAGCTCCATATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 380
 QY 381 TCCCTTCAAGGAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 440
 DB 381 TCCCTTCAAGGAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 440
 QY 441 GCTGCGAAGGAATTCACCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 500
 DB 441 GCTGCGAAGGAATTCACCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 500
 QY 501 ATATTAATCTGTCAGAGAGCTGAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 560
 DB 501 ATATTAATCTGTCAGAGAGCTGAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 560
 QY 561 ATTTGATTTATGAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 620
 DB 561 ATTTGATTTATGAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 620
 QY 550 -TTATGTCATGAGAGAGCTCCATATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 606
 DB 607 CCATCTGAGATGAGAGAGCTCCATATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 662

RESULT 4
 BG722815 775 bp mRNA linear EST 08-MAY-2001
 LOCUS 602695278F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4827427 5',
 DEFINITION mRNA sequence.

ACCESSION BG722815 GI:14002002
 VERSION BG722815
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Butelostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 775)
 NIH-MGC http://imgc.ncl.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-r@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shihrai
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLNL0743 row: 1 column: 20

High quality sequence stop: 651.
 Location/Qualifiers

FEATURES
 source 1..775
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:4827427"

```

/lab host="DH10B"
/clone lib="NIH MGC 97"
/notes: Toram: testis; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed primer 5'-TTTTTTTTT-3',
size-selected for average insert size 2.2 kb and
normalized to 10^5. This is a primary library enriched
for full-length clones and in preparation. Library
Cap-trapped method (Carntest, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH MGC Library.
BASE COUNT      191 a      212 c      193 g      179 t
ORIGIN
Query Match      24.4%; Score 345.4; DB 10; Length 775;
Best Local Similarity 75.1%; Pred. No. 1.5e-67;
Matches 501; Conservative 0; Mismatch 151; Indels 15; Gaps 5;

21 GGGGGGCTGGGAGATGAGTGAACAGAGCCAGAGCCGTCGTCGGGCGCTC 80
39 GAGGCGCCGAGAGATGCGGCGTGAATCAGAGCCAGAGCCGCGGAGCCCTC 98
81 ATCCCTCTGGCGAAGATGTCCTTGAAGCATGTGAGATGTGAGCTCTGCTTCA 140
99 ATCCCTAAGGATGAAGTCTCTGAAGCGGCTCCGATGTGAGCTCTCTCCAG 158
141 AAGCAGTGAATCTATCTCTTTAATGCAAGAAAGAAAGCTTTGCTCACTCA 200
159 CGATCGAAGAGCTCAATGTCTTATGTGAGAAAGACAGAAATGTTCTCACTTCA 218
201 TACCGGATGCTCTGATGCTCACTTATGATGATGATGATGATGATGATGATG 260
219 TACCGGATGCTCTGATGCTCACTTATGATGATGATGATGATGATGATGATG 278
261 CCCTTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 320
279 CATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 338
321 AAGGAAACATTCAGAGAGCTCCATGATGTGTGCGGAGAGAGAGAGAGAGAG 380
339 AAGGAAACATTCAGAGAGCTCCATGATGTGTGCGGAGAGAGAGAGAGAGAG 398
381 TCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 440
399 GTCTTGAAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458
441 GCTGCAAGAGAGATTCATCTGGAAGTGAAT-TTACTGGTTCAGACTTCAGAG 499
459 GCTGCGCCGAGAGATTCATCTGGAAGTGAAT-TTACTGGTTCAGACTTCAGAG 518
500 CATTAATTAATCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 559
519 TGTAAATTAATCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 568
560 AATTAATTAATCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 618
569 --TTATGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 626
619 CTCACCAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 678
627 C-CCACTGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 685
679 GAGGCGCC 685
686 GAGGCTC 692

```

```

RESULT 5
LOCUS      BY715414      717 bp      mRNA      1linear      EST 17-DEC-2002
DEFINITION BY715414 RIKEN full-length enriched, adult male testis Mus musculus
            CDNA clone 4930521123 5', mRNA sequence.
ACCESSION  BY715414

```

VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS

BY715414.1 GI:27128531
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 717)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Onato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H.,
Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A.,
Quackenbush, J., Schreml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,
Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Choila, C., Corbani,
L.B., Cousine, S., Dalla, E., Dragani, T.A., Fletcher, C.P., Forrest,
A., Frazer, K.S., Gasterland, T., Garibol, M., Gissi, C., Godzik, A.,
Gough, J., Grifmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,
Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedler, R.M.,
Kling, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,
P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki,
H., Nagashima, T., Numata, K., Okido, T., Pavan, M.J., Pertea, G.,
Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.V., Qi, D.,
Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring,
B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C.A., Setou,
M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale,
R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y.,
Watanabe, Y., Wells, C., Wilming, L.G., Wyshak-Boris, A., Yanagisawa,
M., Yang, T., Yang, L., Yuan, Z., Zavalon, M., Zhu, Y., Zimmer, A.,
Carninci, P., Hayatsu, N., Hirozane-Kobayashi, T., Kono, H., Nakamura,
M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, M., Imotani, K., Ishii,
Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,
K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,
E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE JOURNAL MEDLINE PUBMED COMMENT

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda,
S., Hashizume, M., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,
Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Kono,
H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,
Numazaki, R., Ohno, M., Ohnishi, N., Saito, R., Sakazume, N., Sano, Y.,
Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y.,
Waki, K., Watanabe, A., Watanabe, M. and Hayashizaki, Y. Direct
Submission

Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multichannel sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES
SOURCE

1. 717
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4930521123"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="SOLR"
/clone_lib="RIKEN full-length enriched, adult male testis"
/note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'}. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 5'
GAGAGAGAGAGCTCGCGCAATTAATTCGATTAATTAATTCCTCCCTCC
-3'}. cDNA was cloned into the XhoI and BamHI sites."

BASE COUNT 161 a 193 c 198 g 165 t

ORIGIN

Query Match 24.2%; Score 341.8; DB 14; Length 717;
Best Local Similarity 70.6%; Pred. No. 9.6e-67;
Matches 506; Conservative 0; Mismatches 192; Indels 19; Gaps 3;

1 CGGCAAGAGGGCGGCGAGAGGGGGCTT-GGGCAGAGATGGCAGTAAACCAAGCCACC 59
19 CGGCGGAGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 78
60 GAGAGCCGCTGTGGGGCCCTCATCCCTCTGGCGAAAGTGTGTTGAAGCAGTGTAGAGT 119
79 GTGAGCCGCGGTGGGGCCGCGCATCCCTCATGTGTGAAGTCTCTTGAAGAAAGTTCAGAA 138
120 GTGAGCTCTGCTCTCTAGAGAAAGAGTGAATCTTCTTTAATGGCAGAAAGAA 179
139 GTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 198
180 GGAAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 239
199 GGGGCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 258
240 CCCATGCTTCTTTTATGATGCGGCTTGGCTGATGATGATGATGATGATGATGATGATGATGAT 299
259 CCCATGCTTCTTTTATGATGCGGCTTGGCTGATGATGATGATGATGATGATGATGATGATGAT 318
300 ATTTTGGCCCAACATTAAGGAACATTTGAGGAGCTCCAGGAGTGGAGTGGAGAA 359
319 ATCTTTGTGCAAACTCAATTAAGGAACATTTGAGGAGCTCCAGGAGTGGAGTGGAGAA 378
360 GGAAGAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 419
379 GGTTCGTCTACTTTTAATATGCTTCGGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 438
420 ATGTGTAAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 479
439 ATGTGTAAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 498
480 TTTCAGACCTTCAGAGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 539
499 ATGTGTAAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 557
540 ACACTTGTGCAAGATATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 599
558 ATATCC-----AAGTTGCTATGAGAGCCCACTGGGAGATATGGGCGC 600

OR 600 CAACAGAGGAGATATGAGATCTCCACAGAGAGATATGAGCCCAACAGAGGAGATATGAG 659
DB 601 TCACCCGTGGAGATACGAGAGTCCATCTGACAGGAGATATGAGCCCAACAGAGGAGATACGA 660
OR 660 GCCCCACCTTATGAGATATGAGAGCCCGCTGTGAGATATGAGATCCCACTGGGGGA 716
DB 661 GCCCGGCGCGGTGGGTATGTATACCCCATCTCCGAGATACATGCTGCTCTCTGGA 717

RESULT 6
EG701881 701 bp mRNA linear EST 07-MAY-2001
LOCUS 602683237F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4815849 5',
DEFINITION mRNA sequence.
ACCESSION EG701881
VERSION EG701881.1 GI:13972665
KEYWORDS EST.
ORGANISM Homo sapiens (human)
SOURCE Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 701)
NIH-MGC <http://mhc.nci.nih.gov/>.
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Mikhail Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki Tohiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10713 row: j column: 10
High quality sequence stop: 680.

FEATURES
SOURCE

1. 701
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4815849"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescript (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (grecag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3', size-selected for average insert size 2.5 kb and normalized to R0.5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC library."

BASE COUNT 176 a 164 c 189 g 172 t

ORIGIN

Query Match 15.0%; Score 211.4; DB 10; Length 701;
Best Local Similarity 67.9%; Pred. No. 2.9e-37;
Matches 464; Conservative 0; Mismatches 191; Indels 28; Gaps 11;

21 GGGGGCTGGGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 80
22 GAGGCCCAAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 80
DB 81 ATCCCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 139
DB 81 ATCCCTTAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 140
OR 140 GAACAGAGATCTCTCTTAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 199
DB 141 GCGATCAGAGAGCTCAATATGCTTTATGTGTAAGAGAGAGAGAGAGAGAGATATCACTTGGC 200

QY 1 CGGACAGAGGGGCGGAGAGGGGCTT-GGGAGAGTGTGAGTAAACGAGGCGACCC 59
 DB 20 CGGCGAGAGCGGCGGCGGAGAGAGAGCTTAAAGCAAGATGACAGTAAACGAGACATACA 79
 QY 60 GAGAGCGGTGTGGGGGCGCTCATCCCTGTGCGAAGTGTCTTGAAGAGTGTAGAT 119
 DB 80 GTGAGACGCGCGGTGTGGGCGGCGCATCTTCATGTGAAAGTCTCTTGAAGAAATGTTCAGAA 139
 QY 120 GTGAGACCTGTGCTTCTTCAAGAAACAGTGAATCTTATCTTTAAATGCGACAAAGAAA 179
 DB 140 GTGAGACCTGTCTTCTTCCCGAGTACCGCAGGCTTCAACCTTTAGTGTGACAAAGAGG 189
 QY 180 GGAACGTGTGTCTTCACTTCAATACCGGTGTCTTGTGTGCTTCACTTCACTTCACTTCACT 239
 DB 200 GGGGCTGTGTCTTCACTTCAATACCGGTGTCTTGTGTGCTTGTGTGCTTGTGTGCTTGTGTG 259
 QY 240 CCATGCTTCTTATGATGATGCGGTGTGCTTGTGTGCTTGTGTGCTTGTGTGCTTGTGTGCT 299
 DB 260 CCATGCTTCTTATGATGATGCGGTGTGCTTGTGTGCTTGTGTGCTTGTGTGCTTGTGTGCT 319
 QY 300 ATTTTGGCCCCCACTTAAAGAAACCAATTCAGGCGAGC 340
 DB 320 ATCTTGTGCAAACTAATTAAAGGAGCAATTCAGGCGAGC 360

RESULT 8
 BU107273 932 bp mRNA linear EST 25-NOV-2002
 LOCUS 60311954F1 CSBQCHL12 Gallus gallus CDNA clone CHEST60K13 5', mRNA
 DEFINITION sequence.

ACCESSION BU107273
 VERSION BU107273.1 GI:25309754
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 932)
 Boardman, P.R., Sans-Beguerro, J., Overton, I.M., Burt, D.W., Bosch, B.,
 Phasianidae; Gallus.
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Bacteria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Chordata; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianidae; Gallus.

TITLE JOURNAL MEDLINE PUBMED
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 10D, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES

Location/Qualifiers
 1..932
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST60K13"
 /dev_stage="36"
 /lab_host="DH10B"
 /note="Organ: heart; Vector: pBluescript II KS(+); Site 1:
 EcoRI, Site 2: NotI; Modification of pBluescript II KS(+)
 [Stratagene] vector to accommodate cDNA produced with the
 T-timed protocol (Construction of uni-directionally
 cloned cDNA libraries from messenger RNA for improved 3'
 end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387
 624). Out pBluescript II KS(+) with NotI and EcoRI.
 Digest in double stranded adaptor containing BglI and
 BamHI sites [5'ggccgcgtcagcccgatccgaaataaag]

BASE COUNT 215 a 273 c 233 g 211 t
 ORIGIN [5'aattcttttttcgataccggggtcgcagc]

Query Match 13.0%; Score 183.6; DB 13; Length 932;
 Best Local Similarity 59.5%; Pred. No. 5.4e-31;
 Matches 350; Conservative 0; Mismatches 229; Indels 9; Gaps 2;

QY 42 GTGAACCAAGCCACACCGAGAGCGTGTGTGGGCGCTCATCCCTCTGGCGAAAGTGTG 101
 DB 21 GCGCTCAATTAAGGACACCTCGAAGAGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 80
 QY 102 TTGAACGATGTGAGATGTGACCTCTGCTTCTTCAAGAAACCAATGTGAATCTATCTC 161
 DB 81 CTGAAGCATGTAAATGTGAGCTCTCTTCAAGATGACATGACGGCAAGCTTGAAGGCC 140
 QY 162 TTTAATGCAAAAGAAAGAAAGCTTTTCTCACTTCAATACGGGTGTCTTGTGACT 221
 DB 141 TTCAAAAGGACCAAGAAAGGAAATGCTGTATCTCAACCCATACAGATGATCTTGTGTG- 199
 QY 222 TCACACTTGTGAATGACCCCATGCTTCTTTTATGATGCGGTGTGGCGTGAATGTGAC 281
 DB 200 -----AAAGGCAAGGATCTTATGCTTGTATATATGCGTTTATTTGTGAAAGG 254
 QY 282 TGACCATTTGAACCAACATTTTGTGCCCCCACTACATTTAAAGAAACATTCAGGAGCT 341
 DB 255 TGCTCTATTGAGAGAGCTCTTCTTCTGTCTATTAATCAATGAAGACATTCAGGCTGAG 314
 QY 342 CGAGGTGTGTGTGAGAGGACAGAGCTTTTAAATTAATCTTCAGAAAGAGGTGCC 401
 DB 315 GCAGAGAGGTGTGTGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 374
 QY 402 ATGGAATTTGCCCACTGATGTGTAAGAGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 461
 DB 375 ATGGAATTTGAGAGAGCTGATGTGTAAGAGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 434
 QY 462 GGAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 521
 DB 435 CAGCCCCCTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 484
 QY 522 GCAATGTGTCTCTCAGACAGACCC--TTGTCCAGCATTCATTCATTCATTCATTCATTCAT 578
 DB 495 CAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 554
 QY 579 CACACCAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 626
 DB 555 GCTCCAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 602

RESULT 9
 BU961805 814 bp mRNA linear EST 21-OCT-2002
 LOCUS AGENCOURT 10617160 NIH_MGC_169 Mus musculus CDNA clone
 DEFINITION IMAGE:6742447 5', mRNA sequence.
 BU961805
 ACCESSION BU961805.1 GI:24191377
 VERSION BU961805.1
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 814)
 NIH-MGC http://mgc.nci.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished
 JOURNAL Contact: Robert Strauberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Jonathan Kuo, NIMH
 CDNA Library Preparation: Michael Brownstein Laboratory
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

http://image.lnl.gov
 Plate: LHC3080 row: 1 column: 06
 High quality sequence stop: 504.
 Location/Qualifiers

FEATURES

Source

1. 814
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6742447"
 /lab_host="DH10B (T1-phage-resistant)"
 /clone_1lb="NH_MGC_169"
 /note="Organ: Testicles; Vector: pDNR-LIB; Site 1: 8fil (ggccatcagcc); Site 2: 5fil (ggccgcgcgcgc); cDNA made by oligo-dt priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
 5'-AAGCAGGATGATACGACGAGGACGACATG-3' and
 5'-ATTCAAGGCGGACGAGGCGGACGATG-3' (30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."
 Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

BASE COUNT 207 a 238 c 177 g 190 t 2 others

ORIGIN

Query Match 12.9%; Score 182.2; DB 13; Length 814;
 Best Local Similarity 65.1%; Pred. No. 1.1e-30;
 Matches 304; Conservative 0; Mismatches 153; Indels 10; Gaps 2;
 615 GGAATCCACAGAAAGATATGAGCCCAAGGAGGATATGAGCCCACTTATGGA 674
 3 GGCTCCCTCTCTCTATATGATGATACCCCAATGGCTATGAGTTCACTCTCTGGA 62
 675 TATGAGCCCGGCTGTGAGATATGAGTCCCACTGGGGATATGAGTCCCACTGGG 734
 63 TATGAGCCCTCACTGTGAGATACGATCCCACTCTGATATGAGCCCACTATAG 122
 735 GATATGAGTCCCACTGGGGATATGAGCCCACTGGGGATATGAGTCCCACT 794
 123 GATATGAGCCCACTCTCAATATGAGAACTCACTATGAGATCCGTTCTCTCACT 182
 795 GGGGATATGAGTCCCACTGGGGATATGAGCCCACTGAGATATGAGCCCA 854
 183 CCCAGATATGAGCCCGGCTATGAGGATATGAGATCTCTTCAAGGGGATAGTCATA 242
 855 CCACTGGAATGAGCCCTTACCCCTGATATGAGTCTCTGTAAGAAATCAAGT 914
 243 CCTCTGATCAAGGGCCAGTCTGTGAGCAAGAGCTCACTGCTGAGCTGAGCA 302
 915 GCTCTCAAGATCTATGACACTGACA-----GGAGATTCTTCCCACTACC 965
 303 GGCATCCCATGTCTGTGAGATGCAATCTGAAATTCAGAGCATTTTCTCTTACC 362
 966 TCATCTTCTAGTCCATTTACCACTTCTGAGAGTTAACTTGAAGATCAACCAAGC- 1024
 363 TGTCTCTCAAGATCACTCTCCGCTTAAAGATTAACCTTTGAAGTTTCAACAGCA 422
 1025 AAAAGGCACTTAAACTGAGTCAAGTAAAGAGAGAGCAAGT 1071
 423 AAAAGGCACTTAAAGTGAAGTCAAGATTAACCCGGAACATTAAGT 469

RESULT 10

BUI03740 560 bp mRNA linear EST 25-NOV-2002
 LOCUS 603005979F1 C8BQCHL01 Gallus gallus CDNA clone CHB8T23n20 5', mRNA
 DEFINITION
 accession BUI03740
 version BUI03740.1 GI:25304305
 keywords EST.
 source Gallus gallus (chicken)
 organism Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Ruteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

REFERENCE

1 (bases 1 to 560)
 Boardman, P.B., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, B.,
 Pong, W.T., Rickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 PUBMED 12445392

COMMENT

Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES

Source

1. 560
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hitec"
 /db_xref="taxon:9031"
 /clone="CHB8T23n20"
 /issue_type="whole embryo"
 /dev_stage="20-21"
 /lab_host="DH10B"
 /clone_1lb="C8BQCHL01"
 /note="Organ: whole embryo; Vector: pBluescript II KS(+);
 Site 1: EcoRI; Site 2: NotI. Modification of pBluescript
 II KS(+) [Stratagene] vector to accommodate cDNA produced
 with the T-primed protocol (construction of
 uni-directionally cloned cDNA libraries from messenger RNA
 for improved 3' end DNA sequencing by Glenn Fu, et al.,
 U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with
 NotI and EcoRI. Ligate in double stranded adaptor
 containing BglI and BamHI sites
 [5'ggcgctgcagccggcgatccggaataaagj
 5'atctcttttcggatccggcgagcgagc]."
 BASE COUNT 131 a 142 c 159 g 128 t

ORIGIN

Query Match 12.7%; Score 179.4; DB 13; Length 560;
 Best Local Similarity 66.1%; Pred. No. 5e-30;
 Matches 277; Conservative 0; Mismatches 136; Indels 6; Gaps 1;
 42 GTGAACGAGCACAACGAGAGCGGTGTGAGGCTCTATCCCTCTGCGAAGTTC 101
 42 GCGCTCAATAGGCAACACTCGAAGAGAGCGGTGTGAGGCTCTATCCGAGATTT 101
 102 TTGAAGCATGTGAGATGTGACTCTGCTTCTCAAGAAACAGTGAATCTATCTC 161
 102 CTCAAGCATGTAAAGATGTGAGCTCTCTCAAGTGAATGACGGCAAGCTTGAAGCC 161
 162 TTTATGCAAGAAAGAAAGAAAGTGTCTCACTTCAATCCGAGTGTCTCTGACT 221
 162 TTCAAGGCAACAAAGAAAGAAAGTGTATCTACCCCAATCAAGATATCTTCTGTCT 220
 222 TCACACTTATGATGACCCCATGCTTCTTATATGATGCGTGTGCTGATGATGAC 281
 221 -----AAAGGCAAGATCTATGCTGTCTTATATGATGCGTGTATTTGTGAAAGG 275
 282 TGACCAATTAACAACAATTTTGCCTCCCAATGATTAAGAAACAATTCAGGCACT 341
 276 TGCTTATTTGAGCAGCTCTTTTCTCTATATATCAACAAAGGCAATTCAGGCTGAG 335
 342 CCAAGTGTGCTGTGAGAGGAGCAAGCTTTTAAATTAATCTTCTGAGAAAGAGATGCC 401
 336 GCAAGAGGTGTGTGAGAGGAGGAGGAAAGTTTAACTGACTTTCAACAGGAGAGAGCC 395
 402 ATCGAATTTGCCCACTGATGATTAAGAGTCTCTGTGCTGCTGCAAGAGATTCACAT 460
 396 ATCGAGTTTGAACAGCTATATTTCAAGAGCTGCTTATGCTTCAAGTGAAGTTCTCT 454

```

RESULT 11
LOCUS      BU284977
DEFINITION BU284977 700 bp mRNA linear EST 27-NOV-2002
            603603379F1 CSEQCEN54 Gallus gallus cDNA clone CHEST581n20 5', mRNA
            sequence.
ACCESSION  BU284977
VERSION    BU284977.1 GI:25734433
KEYWORDS   EST.
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 700)
            Boardman,P.B., Sanz-Bzquez,J., Overton,I.M., Burr,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
            A Comprehensive Collection of Chicken CDNA
            Curr. Biol. 12 (22), 1965-1969 (2002)
JOURNAL    22335534
MEDLINE    12445392
PUBMED
COMMENT
            Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.

FEATURES
            source
            1..700
                /organism="Gallus gallus"
                /mol_type="mRNA"
                /strain="Compton line 151"
                /db_xref="taxon:9031"
                /clone="CHEST581n20"
                /sex="Female"
                /tissue_type="not cerebrum or cerebellum"
                /dev_stage="adult"
                /lab_host="DH10B"
                /clone_id="CSEQCEN54"
                /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
                EcoRI; Site_2: NotI; This normalized library was
                constructed from 1 million independent clones. cDNA
                synthesis was initiated using an oligo(dt) primer, using
                methylated C in the first strand synthesis reaction.
                Following this first strand reaction, double-stranded cDNA
                was bluntend, ligated to NotI adapter, digested with EcoRI
                , size-selected, and cloned into the NotI and EcoRI
                compatible sites of a custom modified MCS of the
                pBluescript (KS+) vector. The library was normalized in 2
                rounds using conditions adapted from Soares et al., PNAS
                (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                (1996): 791, except that a significantly longer
                reannealing hybridization was used."

BASE COUNT      162 a 196 c 178 g 164 t
ORIGIN
Query Match      12.7%; Score 179.4; DB 13; Length 700;
Best Local Similarity 66.1%; Pred. No. 4.9e-30;
Matches 277; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY      42 GTGAACCAAGCCACACCGAGCCGTCGTGGGGCCCTCATCCCTCTGCGAAAGTGC 101
DB      6 GCGCTCAATAGGCACCACTCGAAGGAAGCGGTGTCGTCGCCCAATGCCAAGACATT 65
QY      102 TTGAAGCAGTGTGAGATGTCCTGCTTCTACGAAACAGTGAATCTTATCTC 161
DB      66 CTCAGCAATGTAAAGATGTGAAGCTCTTCCTTCAGTGCATGACGAGGCAACCTTGAAGCC 125
QY      162 TTATATGCAACAAGAAAGAAAGCTTTTCTCATCTTCAATCCGGGTGATCTTGCTGACT 221
DB      126 TTCAAGGCAACCAAGGAAGGAATGCTATCTCAACCCCATACAGGATGATCTTGTGTCTC- 184

```

```

QY      222 TCACACTTAGTCGAATGACCCCAATGCTTTCTTTATGATGCCGTTGGCCGTGATGATGAC 281
DB      185 -----AAAGGCAAGATCTCTATGCTGCTTTATATATGCGGTTATTTGTGAAGG 239
QY      282 TGCACCATTTGAAACCAATTTTGGCCCACTAATTAAGAAACCAATTCAGGAGCT 341
DB      240 TGCCTATATGAGAGGCTCTTTCTCTGCTAATATCAATCAAGACAGATTCAGGCTGAG 299
QY      342 CCAAGTGTGCTGCGAAGGACCAAGCTGTTTAACTTATCTTCAAGAAAGAGTGC 401
DB      300 GCAGAGATGTGCTGCGAAGGCGAGGACGTTTAACTGACTTCAACAGCGAGAGGCC 359
QY      402 ATGCAATTTGCCCACTGATGTAAGAGTGCCTGCTGCTGCCAGAGAAATTCACCT 460
DB      360 ATGAGATTGGAACAGCTGATGTTCAAGAGCTGCTATGATGCTTCAGATGAGATTCCTCT 418

RESULT 12
LOCUS      AJ452591
DEFINITION AJ452591 723 bp mRNA linear EST 22-APR-2002
            AJ452591 riken1 Gallus gallus cDNA clone 31her1, mRNA sequence.
ACCESSION  AJ452591
VERSION    AJ452591.1 GI:20262687
KEYWORDS   EST.
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 723)
            Buerstedde,J.M.
            Gallus gallus Bursal lymphocyte EST
            Unpublished
            Contact: Buerstedde JM
            Cellular Immunology
            Heinrich-Pette-Institute
            Martinstr. 52, 20251 Hamburg, Germany
            Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.

FEATURES
            source
            1..723
                /organism="Gallus gallus"
                /mol_type="mRNA"
                /db_xref="taxon:9031"
                /clone="31her1"
                /cell_type="Bursal lymphocyte"
                /dev_stage="2-3 weeks old"
                /clone_id="riken1"
                /note="CB indred strain"

BASE COUNT      161 a 205 c 181 g 172 t 4 others
ORIGIN
Query Match      12.7%; Score 179.4; DB 9; Length 723;
Best Local Similarity 66.1%; Pred. No. 4.9e-30;
Matches 277; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY      42 GTGAACCAAGCCACACCGAGCCGTCGTGGGGCCCTCATCCCTCTGCGAAAGTGC 101
DB      15 GCGCTCAATAGGCACCACTCGAAGGAAGCGGTGTCGTCGCCCAATGCCAAGACATT 74
QY      102 TTGAAGCAGTGTGAGATGTCCTGCTTCTACGAAACAGTGAATCTTATCTC 161
DB      75 CTCAGCAATGTAAAGATGTGAAGCTCTTCCTTCAGTGCATGACGAGGCAACCTTGAAGCC 134
QY      162 TTATATGCAACAAGAAAGAAAGCTTTTCTCATCTTCAATCCGGGTGATCTTGCTGACT 221
DB      135 TTCAAGGCAACCAAGGAAGGAATGCTATCTCAACCCCATACAGGATGATCTTGTGTCTC- 193
QY      222 TCACACTTAGTCGAATGACCCCAATGCTTTCTTTATGATGCCGTTGGCCGTGATGATGAC 281
DB      194 -----AAAGGCAAGATCTCTATGCTGCTTTATATGATGCGGTTTATTTGGGAAGGG 248
QY      282 TGACACATTGAACAACAATTTTGGCCCACTACATTAAAGAACCAATTCAGGAGCT 341

```

Db 249 TGCCTATTGAGCAGCTCTTTCTCTGTATATCATTAAGAGACAGATTTCAGCTTAG 308
 Qy 342 CCAAGTGTGCTGTGGAGAGACAGCTGTTTAAATATCTTCAGAGAGAGTGC 401
 Db 309 GCAGAGGAGGCTGGAGAGGAGGAGAGCTTTAACTGATTTCAACGAGAGAGG 368
 Qy 402 ATCGAATTTGCCAATCTGATGTGTAAGAGCTGCTGCTGCCAGAGAGAAATTCAC 460
 Db 369 ATCGAGTTTGACACGCTGATGTTCAAGCTGCTTAGTGTCCAGTGGAGTTCTCT 427
 RESULT 13
 B1390687 568 bp mRNA linear EST 06-AUG-2001
 LOCUS B1390687
 DEFINITION Pp2n.pk001.p22 Normalized Chicken Pituitary/Hypothalamus/Pineal
 Library Gallus gallus cDNA clone Pp2n.pk001.p22 5' similar to
 gbl|A010951.1 (U79458) WW domain binding protein-2 (Homo sapiens),
 mRNA sequence.
 ACCESSION B1390687
 VERSION B1390687.1 GI:15083969
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 568)
 AUTHORS Porter,T.E. and Cogburn,L.A.
 TITLE ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
 library USDA/IRPS Animal Genome Project
 JOURNAL Unpublished
 COMMENT Contact: Larry A. Cogburn
 University of Delaware
 Townsend Hall, Newark, DE 19717, USA
 Tel: 302-831-1335
 Fax: 302-831-2822
 Email: cogburn@udel.edu, www.chickest.udel.edu.
 FEATURES
 source location/Qualifiers
 1..568
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Commercial broiler chicken"
 /db_xref="taxon:9031"
 /clone="Pp2n.pk001.p22"
 /sex="Male and Female"
 /tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
 /dev_stage="Embryonic (dl2,dl4,dl9); post-hatch (w1,w3,w5
 ,w7,w9)"
 /lab_host="B. coli EMDH10B"
 /clone_lib="Normalized Chicken
 Pituitary/Hypothalamus/Pineal Library"
 /note="Vector: pCMVSPORT6; Library made from equivalent
 pools of total RNA isolated from each tissue at different
 ages. Single pass sequencing from 5'-end"
 BASE COUNT 128 a 146 c 159 g 130 t 5 others
 ORIGIN
 Query Match 12.6%; Score 177.8; DB 12; Length 568;
 Best Local Similarity 65.9%; Fred. No. 1.1e-29;
 Matches 276; Conservative 0; Mismatches 137; Indels 6; Gaps 1;
 Qy 42 GTGAACGAGCCACACGAGAGCGGTGAGGCGCTCATCCCTCGAGAAAGTGC 101
 Db 26 GCGCTCATATGAGCACATCGAAGAGCGGTGCTGTCTCCCAATGCCAGAGCAT 85
 Qy 102 TTGAAGCATGTGAGATGTGACCTTGCTTCTTACAGAAACAGTGAATCTATCTC 161
 Db 86 CTCAAGCATATGATGATGTGAGCTCTCTTCATGATGATGACGGCAAGCTTAGAGCC 145
 Qy 162 TTTATATGCAAAAGAAAGAAAGCTTTTCTCATCTTACACCGAGTGTCTTCTGACT 221
 Db 146 TTCAAGGACCAAGAGAGGAAATGCTGATCTCACCCCATACAGATGATCTTCTGTC- 204

Qy 222 TCACACTTATGCAATGACCCCATGCTTCTTTATGATGCGGTTGGCTGATGATGAC 281
 Db 205 -----AAGGGGAGAGATCTTATGCTGCTTTTATGAGGCAATTTATTTGTGAAGGG 259
 Qy 282 TGCAACATTTGAACAACAAATTTTGGCCCAACTATTAAGAAACAAATTCAGGACACT 341
 Db 260 TGCCTATTGAGCAGCCCTCTTTCTCTGTATATCATTAAGAGACAAATTCAGGCTAG 319
 Qy 342 CCAAGTGTGCTGTGGAGAGACAGCTGTTTAAATGTTATCTTCAGAGAAAGAGTGC 401
 Db 320 GCAGAGGTGCTGTGGAGAGGAGGAGAGAGCTTTAACTGATTTCAACGAGAGAGAGCC 379
 Qy 402 ATCGAATTTGCCAATCTGATGTGTAAGAGCTGCTGCTGCCAGAGAAATTCAC 460
 Db 360 ATCGAGTTTGACACGCTGATGTTCAAGCTGCTTAGTGTCTTCCAGTGGAGTTCTCT 438
 RESULT 14
 B490729 636 bp mRNA linear EST 07-FEB-2002
 LOCUS B490729
 DEFINITION Pp2n.pk004.e6 Normalized Chicken Pituitary/Hypothalamus/Pineal
 Library (Pp2n) Gallus gallus cDNA clone Pp2n.pk004.e6 5' similar
 to gbl|A010951.1 (U79458) WW domain binding protein-2 (Homo sapiens
), mRNA sequence.
 ACCESSION B490729
 VERSION B490729.1 GI:18611660
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 636)
 AUTHORS Porter,T.E. and Cogburn,L.A.
 TITLE ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
 library, USDA/IRPS Animal Genome Project
 JOURNAL Unpublished
 COMMENT Contact: Larry A. Cogburn
 University of Delaware
 Townsend Hall, Newark, DE 19717, USA
 Tel: 302-831-1335
 Fax: 302-831-2822
 Email: cogburn@udel.edu, www.chickest.udel.edu.
 FEATURES
 source location/Qualifiers
 1..636
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Commercial broiler chickens"
 /db_xref="taxon:9031"
 /clone="Pp2n.pk004.e6"
 /sex="Male and Female"
 /tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
 /dev_stage="Embryonic (dl2,dl4,dl9); post-hatch (1,3,5,7,9
 weeks)"
 /lab_host="B. coli EMDH10B"
 /clone_lib="Normalized Chicken
 Pituitary/Hypothalamus/Pineal Library (pp2n)"
 /note="Vector: pCMVSPORT6; Library made from equivalent
 pools of total RNA isolated from each tissue at different
 ages. Single pass sequencing from 5'-end"
 BASE COUNT 143 a 174 c 173 g 146 t
 ORIGIN
 Query Match 12.6%; Score 177.8; DB 12; Length 636;
 Best Local Similarity 65.9%; Fred. No. 1.1e-29;
 Matches 276; Conservative 0; Mismatches 137; Indels 6; Gaps 1;
 Qy 42 GTGAACGAGCCACACGAGAGCGGTGAGGCGCTCATCCCTCGAGAAAGTGC 101
 Db 26 GCGCTCATATGAGCACATCGAAGAGCGGTGCTGTCTCCCAATGCCAGAGCAT 85
 Qy 102 TTGAAGCATGTGAGATGTGACCTTGCTTCTTACAGAAACAGTGAATCTATCTC 161

```
Db      86 CTGAAGCAATGTAAGATGTGAGCTCTCTTCAAGTACATGAGGGCAAGCTTGAGGCC 145
Qy      162 TTTTAATGGCAAAAGAAAGAAAGCTGTTTCTCACTTCATACCGGGTGTCTTGTGACT 221
Db      146 TTCAAAGGACCAAGAAAGGGAATCTGTATCTCAACCCATACAGAAATGATCTGTGTC- 204
Qy      222 TCACACTTAAGTCAATGACCCCATGTCTTTTATGATGAGCGGTGGCGTGAAGTAC 281
Db      205 -----AAAGGCAAGATCTTATGCTGTCTTTATGATGACATTTTATTTGTGAAGGG 259
Qy      282 TGCAACATTTGCAACCAATTTTGGCCCACTACATTAAGAACCATTCAGGCACT 341
Db      260 TGCTCTATTGAGAGCGCTCTTTCTGTCTAATTACATCAAGAGACAGATTCAAGGCTGAG 319
Qy      342 CCAGGTGTGGCTGGGAAGCAAGCTGTTTAAAGTTATCTTCAGGAAGAGGTGCC 401
Db      320 GCAGAGGTGTGCTGGGAAGGCGAGGAAAGCTTTAACTGACTTTCAACAGGAGGAGGCC 379
Qy      402 ATCGAATTTGCCCACTGATGTAAAGCTGCTGTGCTGCGAGAGAAATTCACCT 460
Db      380 ATCGAGTTTGACAGCTGATGTTCAAAGCTGCTGTGAGCTTCAAGTGAAGTCTCT 438
```

RESULT 15

```
AJ454950      772 bp      mRNA      linear      EST 22-APR-2002
LOCUS      AJ454950 riken1 Gallus gallus cDNA clone 4c7r1, mRNA sequence.
ACCESSION      AJ454950
VERSION      AJ454950.1 GI:20265046
KEYWORDS      EST.
SOURCE      Gallus gallus (chicken)
ORGANISM      Gallus gallus
```

```
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
TITLE      Phasianidae; Gallus.
JOURNAL      1 (bases 1 to 772)
COMMENT      Buerstedde, J.M.
          Gallus gallus bursal lymphocyte EST
          Unpublished
          Contact: Buerstedde JM
          Cellular Immunology
          Heinrich-Pette-Institute
          Martinistr. 52, 20251 Hamburg, Germany
          Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
          Location/Qualifiers
```

```
FEATURES
    source
        1..772
            /organism="Gallus gallus"
            /mol_type="mRNA"
            /db_xref="taxon:9031"
            /clone="4c7r1"
            /cell_type="bursal lymphocyte"
            /dev_stage="2-3 weeks old"
            /clone_id="riken1"
            /note="CB inbred strain"
```

```
BASE COUNT      176 a      154 c      210 g      232 t
ORIGIN
```

```
Query Match      11.8%; Score 167.4; DB 9; Length 772;
Best Local Similarity 65.1%; Pred. No. 2.5e-27;
Matches 265; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

Db      42 GTGAACCAAGCCACACCGAGAGCGGTGTGGGCCCTCATCCCTCTGTGGGAAGTGTG 101
Qy      23 GCGCTCAATAGGCACTCGAAGAAAGCGGTGTGTCGCCCAATGCGAGAGCAATT 82
Db      102 TTGAAGAGTGTGAGATGTGACCTGTGCTTCTACAGAAACAGTGAATCTATCTC 161
Qy      83 CTCAGCAATGTAAAGATGTGAGCTCTCTTCAGTACATGATGAGGCAAGCTTGAGGCC 142
Qy      162 TTTTAATGGCAAAAGAAAGAAAGCTGTTTCTCACTTCATACCGGGTGTCTTGTGACT 221
Db      143 TTCAAAGGACCAAGAAAGGGAATGCTGTATCTCAACCCATACAGAGATGATCTGTGTC- 201
```

```
Qy      222 TCACACTTAAGTCAATGACCCCATGTCTTTTATGATGAGCGGTGGCGTGAAGTAC 281
Db      202 -----AAAGGCAAGATCTTATGCTGTCTTTTATGATGACATTTTATTTGTGAAGGG 256
Qy      282 TGCAACATTTGCAACCAATTTTGGCCCACTACATTAAGAACCATTCAGGCACT 341
Db      257 TGCTCTATTGAGAGCGCTCTTTCTGTCTAATTACATCAAGAGACAGATTCAGGCTGAG 316
Qy      342 CCAGGTGTGGCTGGGAAGCAAGCTGTTTAAAGTTATCTTCAGGAAGAGGTGCC 401
Db      317 GCAGAGGTGTGCTGGGAAGGCGAGGAAAGCTTTAACTGACTTTCAACAGGAGGAGGCC 376
Qy      402 ATCGAATTTGCCCACTGATGTAAAGCTGCTGTGCTGCGAGAGAAATTCACCT 448
Db      377 ATCGAGTTTGACAGCTGATGTTCAAAGCTGCTGTGAGCTTCAAGTGAAGAG 423
```

Search completed: December 16, 2003, 17:31:37
Job time : 2917.96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 05:59:34 ; Search time 339.84 Seconds

(without alignments)
10999.596 Million cell updates/sec

Title: US-09-864-291-4_COPY_36_933

Perfect score: 898
Sequence: 1 atggcagctgacacagagcca.....tgctctcacagatcatga 898

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_ph:*
7: gb_pl:*
8: gb_pl:*
9: gb_pl:*
10: gb_ro:*
11: gb_ro:*
12: gb_ro:*
13: gb_ro:*
14: gb_ro:*
15: gb_ro:*
16: gb_ro:*
17: gb_ro:*
18: gb_ro:*
19: gb_ro:*
20: gb_ro:*
21: gb_ro:*
22: gb_ro:*
23: gb_ro:*
24: gb_ro:*
25: gb_ro:*
26: gb_ro:*
27: gb_ro:*
28: gb_ro:*
29: gb_ro:*
30: gb_ro:*
31: gb_ro:*
32: gb_ro:*
33: gb_ro:*
34: gb_ro:*
35: gb_ro:*
36: gb_ro:*
37: gb_ro:*
38: gb_ro:*
39: gb_ro:*
40: gb_ro:*
41: gb_ro:*
```

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	898	100.0	1413	6 AX359658
2	524.4	58.4	2267	9 BC022546
3	519.2	57.8	1266	9 BC022549
4	374.4	41.7	1001	6 AX359665
5	237	26.4	179222	2 BX296515
6	237	26.4	222469	2 BX470149
7	223.2	24.9	129624	2 BX296540
8	220	24.5	220895	2 HS250D10
9	156.8	17.5	168425	2 AC113593
10	156.8	17.5	224086	10 AC104325
11	151.8	16.9	253149	2 AC107527
12	151.8	16.9	270171	2 AC132969
13	128.4	14.3	793	10 AF439026
14	128.4	14.3	1802	10 MP040826
15	122.8	13.7	222469	2 BX470149
16	122	13.6	1691	9 HSU79458
17	122	13.6	1820	9 BC007452
18	122	13.6	1885	6 AR220846
19	122	13.6	1912	6 BC010616
20	121.6	13.5	2170	9 AB072784
21	120.4	13.4	1835	9 AK057881
22	116	12.9	1205	5 BC053232
23	116	12.9	95108	2 AC127818
24	110.2	12.3	106438	9 AL365445
25	106.4	11.8	2065	14 AF410847
26	106.4	11.8	148418	9 AC020698
27	103.8	11.6	454	6 BD027759
28	102.8	11.4	246814	2 AC103482
29	100	11.1	235381	2 AC103031
30	96	10.7	247947	2 AC098163
31	95.8	10.7	110000	2 AC118467_0
32	93.8	10.4	216292	2 AC100927
33	91.6	10.2	239972	2 AC095835
34	91.6	10.2	241519	2 AC097689
35	91.6	10.2	250902	2 AC131965
36	91.6	10.2	251746	2 AC131411
37	91.4	10.2	208880	2 AC128103
38	90.6	10.1	230569	2 AC108250
39	90.6	10.1	238377	2 AC121702
40	90.4	10.1	210680	2 AC123724
41	90.4	10.1	212334	2 AC023070
42	90.4	10.1	315390	2 AC124097
43	90	10.0	262967	2 AC110392
44	89.8	10.0	261914	2 AC122581
45	88.8	9.9	88716	2 AC098644

ALIGNMENTS

RESULT 1
AX359658
LOCUS AX359658 1413 bp DNA
DEFINITION Sequence 4 from Patent WO0190185.
ACCESSION AX359658
VERSION AX359658.1 GI:18675409
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
unclassified.

REFERENCES

1
Okamoto, R. and Sutoy, P.
P32 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses
thereof
Patent: WO 0190185-A 4 29-NOV-2001;
JOURNAL

PAGSARPHBSTAQA PENBASLPASASSOVHS"
 BASE COUNT 709 a 477 c 469 g 612 t
 ORIGIN

Query Match 58.4%; Score 524.4; DB 9; Length 2267;
 Best Local Similarity 76.1%; Pred. No. 1.2e-120;
 Matches 674; Conservative 0; Mismatches 196; Indels 12; Gaps 2;

```

QY 1 ATGGAGTGAACAGAGCCACCGAGCCGCTGCGGGCCCTCATCCCTCGGCGAA 60
DB 32 ATGGCGGTGAATCAGAGCCACCGAGAACCGCGGAGCCCTCATCCCTTAACGTGAA 91
QY 61 AGTGTCTTGAACAGTGTGAGATGTGACCTCTGCTCTCTCAAGAAACAGTGAATCC 120
DB 92 AGTCTCTTGAAGGGGTCTCGAATGTGAGGCTCTCTTCCACAGAGATCAAGAGCTCA 151
QY 121 TATCTCTTAAATGCAAGAAAGAAAGAGAGCTTTCTCATCTTCATACCGGGTGTCTTC 180
DB 152 AATGTCTTAAATGCAAGAAAGAAAGAGAGAGCTTTCTCATCTTCATACCGGGTGAATTTTC 211
QY 181 GTGACTTCAACCTTGAATGACCCGAGCTTTCTTTATATAGTCCGTTTGGCCCTGATG 240
DB 212 ATTAAGCTTCAATGCTCCATGATGATCCAGTGTGCTTTTATATATGCAATTTATGATG 271
QY 241 AGTGAATGCAACATTTGAAACCAATTTTGGCCCACTTCAATTAAGAAACATTTAG 300
DB 272 AGGAACCTCACTGTGAAACCAAGATTTGCTGCAAACTTCATTAAGGAATTAATTCAG 331
QY 301 GAGCTTCAGAGTGTGCTGAGAAAGAGAGCTGTTTAAAGTTATCTTCAGAAAGAA 360
DB 332 GAGGCTCCATATATGTGTGCTGAGAAAGAGAGCTTAAATTAAGTTCTTCAGAAATGGA 391
QY 361 GGTGCTCATGAAATTTGGCCCACTGATGTAAAGCTGCTGCTGCTGCTGCGAGAGAT 420
DB 392 GGTGCTCATGAAATTTGGCCCACTGATGTAAAGCTGCTGCTGCTGCTGCGAGAGAT 451
QY 421 CCACTTGAAGTGAATTAATTAAGTTCAGACCTTCAGAGCTGATTAATTAAGTCCCA 480
DB 452 CCACTTGAAGTGAATTAATTAAGTTCAGAGCTTCAAGTATTAATTAAGTTCAGAGAA 511
QY 481 GGGGCTGAGTGTGCTCTCAGAGAGAGCTTGTCCAGATTAATTAAGTTCAGAGAA 540
DB 512 GGGGAT---ATGTGAGCTCAGAGAGAGCTTGTTCAG-----TTATGTGTATGGA 559
QY 541 CCCCCCAACAGATATTAAGTCCAAAGAGAGAGATTAAGAACTTCACAGAGAGAT 600
DB 560 GCCCCCACTGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 619
QY 601 GAGAGCCCAACAGAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 620 GAGAGCCCAACAGAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 679
QY 661 TATGAGATTCAGAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 680 TATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 739
QY 721 GATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 740 GATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 799
QY 781 GGGGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 800 CTGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 859
QY 841 CCCCCCTGATATGAAGCTCATCTGTGAGAAATTAAGAGAGAGAGAGAGAGAGAG 882
DB 860 CCGCTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901

```

RESULT 3
 BC022549 2266 bp mRNA linear PRI 04-FEB-2002
 LOCUS BC022549
 DEFINITION Homo sapiens, similar to RIKEN cDNA 4930521I23 gene, clone

ACCSSION MGC:26828 IMAGE:4815849, mRNA, complete cds.

VERSION BC022549.1 GI:18490710

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Bukarykova; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 2266)

Strausberg, R.

Submitted (01-FEB-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcdpaxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>

Series: IRAX Plate: 32 Row: n Column: 3

This clone was selected for full length sequencing because it

passed the following selection criteria: Similarity but not

identity to protein.

Location/Qualifiers

1..2266

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:26828 IMAGE:4815849"

/tissue_type="Brain, hippocampus"

/clone_id="NIH MGC_95"

/lab_host="DH10B"

/note="vector: pBluescript"

32..634

/codon_start=1

/product="Similar to RIKEN cDNA 4930521I23 gene"

/protein_id="AAH22549.1"

/db_xref="GI:18490711"

FEATURES

source

CDS

BASE COUNT

ORIGIN

Query Match 57.8%; Score 519.2; DB 9; Length 2267;
 Best Local Similarity 76.1%; Pred. No. 2.4e-119;
 Matches 671; Conservative 0; Mismatches 196; Indels 13; Gaps 2;

```

QY 1 ATGGAGTGAACAGAGCCACCGAGCCGCTGCGGGCCCTCATCCCTCGGCGAA 60
DB 32 ATGGCGGTGAATCAGAGCCACCGAGAACCGCGGAGCCCTCATCCCTTAACGTGAA 91
QY 61 AGTGTCTTGAACAGTGTGAGATGTGACCTCTGCTCTCTCAAGAAACAGTGAATCC 120
DB 92 AGTCTCTTGAAGGGGTCTCGAATGTGAGGCTCTCTTCCACAGAGATCAAGAGCTCA 151
QY 121 TATCTCTTAAATGCAAGAAAGAAAGAGAGCTTTCTCATCTTCATACCGGGTGTCTTC 180
DB 152 AATGTCTTAAATGCAAGAAAGAAAGAGAGAGCTTTCTCATCTTCATACCGGGTGAATTTTC 211

```

QY 181 GAGCTTCAGACTTAACTGATGACCCAGCTTCTTTATATGATGCGCTTGGCTGANG 240
 DB 212 ATATGATGATGCTTCACTGATGATGATGATGATGATGATGATGATGATGATGATG 271
 QY 241 AGTACCTGACCAATGAAACCAATTTTGGCCCACTACATTAAGGACCATTCAG 300
 DB 272 ACAGACTCTGCTTGAACCAAGATTTTGGCTGCAACTCTTAAAGGACCATTCAG 331
 QY 301 GAGCTTCAGAGTGTGCTGGAAGACAGAGCTGTTTAAAGTTATCTTCAAGAAAGA 360
 DB 332 GAGCTTCATATGATGCTGGAAGACAGAGCTTAAATTAATTAATTAATTAATTA 391
 QY 361 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 420
 DB 392 GATGCTTCATATGATGCTGGAAGACAGAGCTTAAATTAATTAATTAATTAATTA 451
 QY 421 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 480
 DB 452 GATGCTTCATATGATGCTGGAAGACAGAGCTTAAATTAATTAATTAATTAATTA 507
 QY 481 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 540
 DB 508 GATGCTTCATATGATGCTGGAAGACAGAGCTTAAATTAATTAATTAATTAATTA 558
 QY 541 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 600
 DB 559 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 618
 QY 601 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 660
 DB 619 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 678
 QY 661 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 720
 DB 679 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 728
 QY 721 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 780
 DB 739 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 798
 QY 781 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 840
 DB 799 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 858
 QY 841 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 882
 DB 859 GATGCTTCAGATTTTGGCCCACTGATGATGATGATGATGATGATGATGATGAT 900
 RESULT 4
 AX359665 1001 bp DNA linear PAT 13-FEB-2002
 LOCUS Sequence 11 from Patent WO0190185.
 DEFINITION AX359665
 ACCESSION AX359665.1 GI:18675411
 VERSION
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 Oke, R. and Sutoy, P.
 P32 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses
 thereof: WO 0190185-A, 11 29-NOV-2001;
 JOURNAL QUBEN'S UNIVERSITY AT KINGSTON (CA); OREGON HEALTH SCIENCES,
 UNIVERSITY (US)
 FEATURES
 SOURCE 1. 1001
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 1. 708

/note="unnamed protein product"
 /codon_start=1
 /protein_id="CAD2337.1"
 /db_xref="GI:18675412"
 /translation="MPDMLTTLTVBQVFPANPFIKGTIOAPYGMBSQATPKLVFR
 NGDAIEPMLWKAASAVRGPPLTLDWFSMDIYYITGGMCTPOMPSTVYG
 APBAGYAPBPBGYAPBAGYAPBPBGYAPBPBGYAPBPBGYAPBPBGYAPBPBGYAP
 PLGAYAPPLGAYAPPLGAYAPPLGAYAPPLGAYAPPLGAYAPPLGAYAPPLGAYAPPLGAYAP
 PBNBSLPBSSQVHS"
 BASE COUNT 261 a 254 c 248 g 238 t
 ORIGIN
 Query Match 41.7%; Score 374.4; DB 6; Length 1001;
 Best Local Similarity 75.3%; Pred. No. 5.1e-83;
 Matches 497; Conservative 0; Mismatches 151; Indels 12; Gaps 2;
 QY 223 ATGCGCTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 282
 DB 1 ATGCGCTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
 QY 283 ATTAAGGAACATTCAGAGCTTCAGAGTGTGCTGGAAGACAGAGCTGTTTAAAG 342
 DB 61 ATTAAGGAACATTCAGAGCTTCAGAGTGTGCTGGAAGACAGAGCTTCTTAA 120
 QY 343 TTAATCTTCAAGGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 402
 DB 121 TTAATCTTCAAGGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
 QY 403 GCTGCTGCAAGGAATTCAGAGTGTGCTGGAAGACAGAGCTTCTTAAAG 462
 DB 181 GCTGCTGCAAGGAATTCAGAGTGTGCTGGAAGACAGAGCTTCTTAAAG 240
 QY 463 TTAATCTTCAAGGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 522
 DB 241 TTAATCTTCAAGGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 292
 QY 523 CCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 582
 DB 293 CCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 348
 QY 583 ACTTCAAGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 642
 DB 349 GCTTCAAGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 408
 QY 643 GAGAGCTTCAAGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 702
 DB 409 GAGAGCTTCAAGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 468
 QY 703 TATGAGTGTGCTGGAAGACAGAGCTTCTTAAAGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 762
 DB 469 TATGAGTGTGCTGGAAGACAGAGCTTCTTAAAGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 528
 QY 763 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 822
 DB 529 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 588
 QY 823 GCTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 882
 DB 589 GCTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 948
 RESULT 5
 BX296515 179222 bp DNA linear HTG 26-MAY-2003
 LOCUS BX296515
 DEFINITION BX296515
 ACCESSION BX296515.6 GI:31076160
 VERSION
 KEYWORDS HTG; HTGS PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Sus scrofa (pig)
 ORGANISM Sus scrofa
 Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 6656: contig of 6656 bp in length
6657 6756: gap of 100 bp
6757 10255: contig of 3499 bp in length
10256 10355: gap of 100 bp
10356 13223: contig of 2868 bp in length
13224 13323: gap of 100 bp
13324 19935: contig of 6612 bp in length
19936 20035: gap of 100 bp
20036 22597: contig of 2562 bp in length
22598 22697: gap of 100 bp
22698 28569: contig of 5872 bp in length
28570 28669: gap of 100 bp
28670 37965: contig of 9296 bp in length
37966 38065: gap of 100 bp
38066 41140: contig of 3075 bp in length
41141 41240: gap of 100 bp
41241 45108: contig of 3868 bp in length
45109 45208: gap of 100 bp
45209 47730: contig of 2522 bp in length
47731 47830: gap of 100 bp
47831 51884: contig of 4054 bp in length
51885 51984: gap of 100 bp
51985 55857: contig of 3873 bp in length
55858 55957: gap of 100 bp
55959 60498: contig of 4541 bp in length
60499 60598: gap of 100 bp
60599 63641: contig of 3043 bp in length
63642 63741: gap of 100 bp
63742 72674: contig of 8933 bp in length
72675 72774: gap of 100 bp
72775 75602: contig of 2828 bp in length
75603 75702: gap of 100 bp
75703 78062: contig of 2360 bp in length
78063 78162: gap of 100 bp
78163 84485: contig of 6323 bp in length
84486 84585: gap of 100 bp
84586 87553: contig of 2968 bp in length
87554 87653: gap of 100 bp
87654 92487: contig of 4834 bp in length
92488 92587: gap of 100 bp
92589 96104: contig of 3517 bp in length
96105 96204: gap of 100 bp
96205 101479: contig of 5275 bp in length
101480 101579: gap of 100 bp
101580 104979: contig of 3400 bp in length
104980 105079: gap of 100 bp
105080 108003: contig of 2924 bp in length
108004 108103: gap of 100 bp
108104 113151: contig of 5048 bp in length
113152 113251: gap of 100 bp
113252 116585: contig of 3334 bp in length
116586 116685: gap of 100 bp
116687 119346: contig of 2661 bp in length
119347 119446: gap of 100 bp
119447 123125: contig of 3679 bp in length
123126 123225: gap of 100 bp
123226 131266: contig of 8041 bp in length
131267 131366: gap of 100 bp
131367 136193: contig of 4827 bp in length
136194 136293: gap of 100 bp
136294 139482: contig of 3189 bp in length
139483 139582: gap of 100 bp
139583 141875: contig of 2293 bp in length
141876 141975: gap of 100 bp
141976 144122: contig of 2147 bp in length
144123 144222: gap of 100 bp
144223 146968: contig of 2746 bp in length
146969 147068: gap of 100 bp

147069 149399: contig of 2331 bp in length
149400 149499: gap of 100 bp
149500 153963: contig of 4464 bp in length
153964 154063: gap of 100 bp
154064 156579: contig of 2516 bp in length
156580 156679: gap of 100 bp
156680 158928: contig of 2249 bp in length
158929 159028: gap of 100 bp
159029 164542: contig of 5514 bp in length
164543 164642: gap of 100 bp
164643 172054: contig of 7412 bp in length
172055 172154: gap of 100 bp
172155 175210: contig of 3056 bp in length
175211 175310: gap of 100 bp
175311 179379: contig of 4059 bp in length
179380 179479: gap of 100 bp
179480 182698: contig of 3219 bp in length
182699 182798: gap of 100 bp
182799 186779: contig of 3981 bp in length
186780 186879: gap of 100 bp
186880 195626: contig of 8747 bp in length
195627 195726: gap of 100 bp
195727 199396: contig of 3670 bp in length
199397 199496: gap of 100 bp
199497 206412: contig of 6916 bp in length
206413 206512: gap of 100 bp
206513 208619: contig of 3107 bp in length
208620 209719: gap of 100 bp
209720 219083: contig of 9364 bp in length
219084 219183: gap of 100 bp
219184 222469: contig of 3286 bp in length.

FEATURES
source
1..222469
Location/Qualifiers
1..222469
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone_lib="DKEY-242K7"
/clone_1ib="DantolKey"
1..6656
/note="assembly_fragment:01359
fragment_chain:1"
6757..10255
/note="assembly_fragment:01385
fragment_chain:1"
10356..13323
/note="assembly_fragment:00592
fragment_chain:1"
13324..19935
/note="assembly_fragment:00237
fragment_chain:1"
20036..22597
/note="assembly_fragment:00949
fragment_chain:1"
22698..28569
/note="assembly_fragment:00615
fragment_chain:1"
28670..37965
/note="assembly_fragment:01306
fragment_chain:2"
38066..41140
/note="assembly_fragment:00487
fragment_chain:2"
41241..45108
/note="assembly_fragment:00232
fragment_chain:2"
45209..47730
/note="assembly_fragment:00935
fragment_chain:2"
47831..51884
/note="assembly_fragment:01415
fragment_chain:2"
51985..55857
/note="assembly_fragment:00878

```

misc_feature      55958..60498
                  /note="assembly: fragment:01208
                  fragment:chain:3"
misc_feature      60599..63641
                  /note="assembly: fragment:01266
                  fragment:chain:3"
misc_feature      63742..72674
                  /note="assembly: fragment:01329
                  fragment:chain:3"
misc_feature      72775..75602
                  /note="assembly: fragment:01183
                  fragment:chain:4"
misc_feature      75703..78062
                  /note="assembly: fragment:01179

Query Match      26.4%; Score 237; DB 2; Length 222469;
Best Local Similarity 73.4%; Pred. No. 1.3e-48;
Matches 303; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 485 CTGCATGCTCTCTCAACAGACCTTGTCCAGCATATCCATTTGATCTATGACCCC 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197009 CTCACAGATACGAGGCCCACTCCAGATACGAGGCCCACTCCAGATACGAGGCC 197068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 545 CACCAACAGATATACAGTCCAAACGAGGGGAATATGAACTCCACCAAGAAAGATATGAG 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197069 CACCTTCAGATACGAGGCCCACTCCAAAGACTCCAAAGGCCCACTCCAGATATGAG 197128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 605 CCCAACAGAGGGGATATGAGGCCCACTATGAGATATGAGGCCCGCTGTGTGATATG 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197129 CCCATCTGGGGATATGAGGCTCACTGCGAGATATGACGCCCTGCTCCAGATATG 197188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 665 GAGTCCACCTGGGGATATGAGGCTCCACTGGGGGAATATGAGTCCCACTGGGGGAT 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197189 GAGCTCACCTGAAGACAGAAAGGCCCAACAGAGATATGAGTCCCATCTGCGGGAT 197248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 725 ATGAGAGGCCCACTGGGGATATGAGGCTCCACTGGGGGAATATGAGTCCCACTGGGG 784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197249 ATGAGAGGCCCACTGGGAATATGAGTCCCGCTCCAGATATGAGGCTCCAGAA 197308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 785 GATATGAGAGGCCCACTGGAGATATGAGGCCCACTGGGAATATGAGGCTCCATCCC 844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197309 GATATGAGAGGCCCTCTGCGAGATATGAGCAACACAGCTGGAATATGAGGCCCACTC 197368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 845 CTGCATATGAGAGCTTCATCTGCGAAATATGAGGCTCTCTCAAGATCTATG 897
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197369 TAGCATATGAGAGGCCCATCTGCGAAATATGAGGCTCTCTCAAGATCTATG 197421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
BX296540/c      129624 bp      DNA      linear      HTG 02-APR-2003
LOCUS           BX296540.2 GI:29500961
DEFINITION      Sus scrofa clone Pigs-231K18, *** SEQUENCING IN PROGRESS ***, 37
                unordered pieces.
ACCESSION       BX296540
VERSION         BX296540.2 GI:29500961
KEYWORDS        HTG; HTGS PHASE1.
SOURCE          Sus scrofa (pig)
ORGANISM        Sus scrofa
                Bacteria; Chordata; Craniata; Vertebrata; Buteleostomi;
                Mammalia; Suidae; Cetartiodactyla; Suidae; Suidae; Sus.
REFERENCE       1 (bases 1 to 129624)
AUTHORS         Burdon, J.
TITLE           Direct Submision
JOURNAL         Submitted (02-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
                Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
                On Apr 2, 2003 this sequence version replaced gi:29335441.
COMMENT         ----- Genome Center
                Center: Wellcome Trust Sanger Institute
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: humquerry@sanger.ac.uk

```

```

----- Project Information
Center project name: br231K18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 113740 bases at least Q40
Consensus quality: 118187 bases at least Q30
Consensus quality: 120712 bases at least Q20
Insert size: 126024; sum-of-contigs
Insert size: 154528; 4.6% error; agarose-fp
Quality coverage: 2.35x in Q20 bases; sum-of-contigs Quality
coverage: 3.19x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 4262: contig of 4262 bp in length
* 4263 4362: gap of 100 bp
* 4363 6391: contig of 2629 bp in length
* 6392 7091: gap of 100 bp
* 7092 9480: contig of 2389 bp in length
* 9481 9580: gap of 100 bp
* 9581 11688: contig of 2108 bp in length
* 11689 11788: gap of 100 bp
* 11789 14658: contig of 2870 bp in length
* 14659 14758: gap of 100 bp
* 14759 18440: contig of 3682 bp in length
* 18441 18540: gap of 100 bp
* 18541 20818: contig of 2278 bp in length
* 20819 20918: gap of 100 bp
* 20919 22969: contig of 2051 bp in length
* 22970 23069: gap of 100 bp
* 23070 27584: contig of 4515 bp in length
* 27585 27684: gap of 100 bp
* 27685 33211: contig of 5527 bp in length
* 33212 33311: gap of 100 bp
* 33312 35521: contig of 2210 bp in length
* 35522 35621: gap of 100 bp
* 35622 38528: contig of 2907 bp in length
* 38529 38628: gap of 100 bp
* 38629 42063: contig of 3435 bp in length
* 42064 42163: gap of 100 bp
* 42164 44164: contig of 2001 bp in length
* 44165 44264: gap of 100 bp
* 44265 47707: contig of 3443 bp in length
* 47708 47807: gap of 100 bp
* 47808 50737: contig of 2930 bp in length
* 50738 50837: gap of 100 bp
* 50838 57066: contig of 6229 bp in length
* 57067 57166: gap of 100 bp
* 57167 59978: contig of 2812 bp in length
* 59979 60078: gap of 100 bp
* 60079 62560: contig of 2482 bp in length
* 62561 62660: gap of 100 bp
* 62661 64863: contig of 2203 bp in length
* 64864 65063: gap of 100 bp
* 65064 67191: contig of 2128 bp in length
* 67192 67291: gap of 100 bp
* 67292 72005: contig of 4714 bp in length
* 72006 72106: gap of 100 bp
* 72106 76595: contig of 4490 bp in length
* 76596 76695: gap of 100 bp
* 76696 79825: contig of 3130 bp in length
* 79826 79925: gap of 100 bp
* 79926 83540: contig of 3615 bp in length
* 83541 83641: gap of 100 bp
* 83641 86276: contig of 2636 bp in length
* 86277 86376: gap of 100 bp

```


ANT2 (adenine nucleotide translocator 2), 2 mRNAs based on ESTs, a genomic marker D22S1178, a CA repeat polymorphism, ESTs and a Cpg island, complete sequence.
 accession 299716
 version 4 GI:4456457
 keywords HTG; ANT2; Cpg Island; D22S1178; NAGA; septin 3; SREBF2.
 source Homo sapiens (human)
 organism Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 220895)
 Clark, G.
 direct submission
 submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 clone request: clonerequest@sanger.ac.uk
 On Mar 21, 1999 this sequence version replaced gi:4164339.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Bm: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C.elegans/wormpep/CTA-250D10 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218.
 VECTOR: pBAC108L.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
 This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr22
 This sequence is the entire insert of clone CTA-250D10. The true left end of clone RPI-18601 is at 129979 in this sequence. The true left end of clone RPI-359J16 is at 1335 in this sequence. The true right end of clone RPI-821D11 is at 23458 in this sequence. The true right end of clone RPI-359J16 is at 118711 in this sequence.
 Location/Qualifiers
 source 1..220895
 /organism="Homo sapiens"
 /mol_type="Genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="22"
 /clone="CTA-250D10"
 /clone_1lb="CT978SK-A1"
 17..29
 /note="2.6 copies 5 mer ATTGG 26% conserved"
 repeat_region 33..325
 /note="AluX repeat: matches 1..301 of consensus"
 repeat_region 326..338
 /note="2.6 copies 5 mer TTGCA 26% conserved"
 repeat_region 667..683
 /note="2.1 copies 8 mer AGAGACC 34% conserved"
 repeat_region 740..751
 /note="2.4 copies 5 mer TTTTA 24% conserved"
 repeat_region 1443..1457

/note="7.5 copies 2 mer AC 21% conserved"
 repeat_region 1715..1997
 /note="AluJ repeat: matches 5..293 of consensus"
 repeat_region 2103..2378
 /note="AluS repeat: matches 2..296 of consensus"
 repeat_region 2844..2855
 /note="12.0 copies 1 mer A 24% conserved"
 repeat_region /note="MIR repeat: matches 70..235 of consensus"
 repeat_region complement(3040..3344)
 /note="AluY repeat: matches 1..310 of consensus"
 repeat_region 3446..3681
 /note="L1M4A repeat: matches 5256..5471 of consensus"
 repeat_region 3746..3837
 /note="L1M4A repeat: matches 5615..5707 of consensus"
 repeat_region 3838..4142
 /note="AluJ repeat: matches 2..305 of consensus"
 repeat_region 4143..4441
 /note="L1M4A repeat: matches 5707..5999 of consensus"
 repeat_region 4442..4747
 /note="AluS repeat: matches 1..292 of consensus"
 repeat_region 4748..4858
 /note="L1M4A repeat: matches 5999..6110 of consensus"
 repeat_region 4881..5172
 /note="AluX repeat: matches 1..296 of consensus"
 repeat_region 5418..5429
 /note="2.4 copies 5 mer GCTGA 24% conserved"
 repeat_region complement(5769..6021)
 /note="match: GGS; Bm:A0069839"
 repeat_region 6074..6089
 /note="2.0 copies 8 mer CCGGCGT 32% conserved"
 repeat_region 6139..6153
 /note="2.5 copies 6 mer TTTTA 21% conserved"
 repeat_region 6141..6155
 /note="3.0 copies 5 mer TTTAA 21% conserved"
 repeat_region complement(6366..6572)
 /note="MIR repeat: matches 5..214 of consensus"
 repeat_region 6839..6852
 /note="2.3 copies 6 mer GAAAT 28% conserved"
 repeat_region 7088..7114
 /note="6.8 copies 4 mer TTTA 36% conserved"
 repeat_region 7091..7109
 /note="1.9 copies 10 mer ATTATTAC 38% conserved"
 repeat_region complement(7122..7418)
 /note="AluX repeat: matches 1..300 of consensus"
 repeat_region 7464..7502
 /note="19.5 copies 2 mer AT 78% conserved"
 repeat_region 7502..7524
 /note="11.5 copies 2 mer AC 46% conserved"
 repeat_region 7523..7538
 /note="2.0 copies 8 mer CAAATAT 32% conserved"
 repeat_region 7535..7563
 /note="14.5 copies 2 mer AT 49% conserved"
 repeat_region complement(7564..7859)
 /note="AluY repeat: matches 1..296 of consensus"
 repeat_region 7873..7893
 /note="5.2 copies 4 mer TGT 33% conserved"
 repeat_region 7914..8200
 /note="AluJ repeat: matches 1..290 of consensus"
 repeat_region complement(8202..8318)
 /note="AluJ/FLM repeat: matches 1..117 of consensus"
 repeat_region 8387..8409
 /note="11.5 copies 2 mer AC 46% conserved"
 repeat_region 8483..8498
 /note="2.0 copies 8 mer GTGGTGT 32% conserved"
 repeat_region 8486..8496
 /note="2.2 copies 5 mer GTGT 22% conserved"
 repeat_region 8606..8622
 /note="2.4 copies 7 mer AGAGAA 34% conserved"
 repeat_region complement(8767..8989)
 /note="MIR115 repeat: matches 33..268 of consensus"
 repeat_region 9004..9014
 /note="2.2 copies 5 mer CTGG 22% conserved"

TITLE
JOURNAL
COMMENT

Vasilev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced g1:21313869.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

Project information

Center project name: L23816

Center clone name: 363_1_24

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 16227 bases at least Q40

Consensus quality: 16485 bases at least Q30

Insert size: 162000; agarose-fp

Insert size: 166725; sum-of-ctnigs

Quality coverage: 8.6 in Q20 bases; agarose-fp

Quality coverage: 8.3 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

1      1412: contig of 1412 bp in length
*
* 1413      1512: gap of 100 bp
*
* 1513      3366: contig of 1854 bp in length
*
* 3367      3466: gap of 100 bp
*
* 3467      5542: contig of 2076 bp in length
*
* 5543      5643: gap of 100 bp
*
* 5643      7949: contig of 2307 bp in length
*
* 7950      8049: gap of 100 bp
*
* 8050      42610: contig of 34561 bp in length
*
* 42611      42710: gap of 100 bp
*
* 42711      45717: contig of 3007 bp in length
*
* 45718      45817: gap of 100 bp
*
* 45818      50106: contig of 4289 bp in length
*
* 50107      50206: gap of 100 bp
*
* 50207      52945: contig of 2739 bp in length
*
* 52946      53045: gap of 100 bp
*
* 53046      56269: contig of 3324 bp in length
*
* 56270      56369: gap of 100 bp
*
* 56370      59897: contig of 3528 bp in length
*
* 59898      59997: gap of 100 bp
*
* 59998      67207: contig of 7210 bp in length
*
* 67208      67307: gap of 100 bp
*
* 67308      78743: contig of 11436 bp in length
*
* 78744      78843: gap of 100 bp
*
* 78844      89991: contig of 11148 bp in length
*
* 89992      90091: gap of 100 bp
*
* 90092      107918: contig of 17827 bp in length
*
* 107919      108018: gap of 100 bp
*
* 108019      124874: contig of 16856 bp in length
*
* 124875      124974: gap of 100 bp
*
* 124975      142093: contig of 17119 bp in length
*
* 142094      142193: gap of 100 bp
*
* 142194      167111: contig of 24918 bp in length
*
* 167112      167211: gap of 100 bp
*
* 167212      168425: contig of 1214 bp in length.
```

FEATURES
source

Location/Qualifiers
1..168425
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-363124"
/clone_1lb="RP23-363124"
1..1412
/note="assembly_fragment"

```

misc_feature      1513..3366  
/note="assembly_fragment"  
misc_feature      3467..5542  
/note="assembly_fragment"  
misc_feature      5643..7949  
/note="assembly_fragment"  
misc_feature      8050..42610  
/note="assembly_fragment"  
misc_feature      42711..45717  
/note="assembly_fragment"  
misc_feature      45818..50106  
/note="assembly_fragment"  
misc_feature      50207..52945  
/note="assembly_fragment"  
misc_feature      53046..56269  
/note="assembly_fragment"  
misc_feature      56370..59897  
/note="assembly_fragment"  
misc_feature      59998..67207  
/note="assembly_fragment"  
misc_feature      67308..78743  
/note="assembly_fragment"  
misc_feature      78844..89991  
/note="assembly_fragment"  
misc_feature      90092..107918  
/note="assembly_fragment"  
misc_feature      108019..124874  
/note="assembly_fragment"  
misc_feature      124975..142093  
/note="assembly_fragment"  
misc_feature      142194..167111  
/note="assembly_fragment"  
misc_feature      167212..168425  
/note="assembly_fragment"
```

BASE COUNT 44391 a 39572 c 38747 g 44015 t 1700 others
ORIGIN

```

Query Match      17.5%; Score 156.8; DB 2; Length 168425;  
Best Local Similarity 66.4%; Pred. No. 1.7e-28;  
Matches 243; Conservative 0; Mismatches 117; Indels 6; Gaps 1;  

QY      529 GTGATCTATGAGACCCCAACGAGATATACATTCACCAAGGAGATATGAACTCCA 588  
DB      66144 GTTCCCTATGAGAGCCCACTCGGGATATAGGGCTCACCCCTGAGATACGAGTCCCA 66203  

QY      589 CCAGAGATATGAGAGCCCAACGAGGAGATATGAGAGCCCACTATGATATGAGAGCC 648  
DB      66204 TCTCAGAGATATGAGAGCCCACTCGGGATATAGAGAGCCCACTGAGATATGAGCC 66263  

QY      649 CCGCTGTGAGATATGAGAGTCCCACTCGGGAGATATGAGATCCCACTCGGGAGATATGAG 708  
DB      66264 CCATCTCCGATACGATGCTCTGCTCTCTGAGATACGAG-----CTGTGAGATATGAGC 66317  

QY      709 GTCCACCTGAGAGATATGAGAGCCCACTGAGGAGATATGAGATCCCACTGAGGAGATAT 768  
DB      66318 TCCCTCTCTCTCTATATGATGCTACCCCAATGAGGCTATGAGATTCACCTCTGAGATAT 66377  

QY      769 GATGAGCCCACTGAGGAGATATGAGAGCCCACTGAGGAGATATGAGAGCCCACTGAGAG 828  
DB      66378 GGAATCTCACTGATGATATGAGATCCCACTCTGAGATATGAGAGCCCACTATGAGAG 66437  

QY      829 AATGAGCCCACTGAGATATGAGAGCTCATCTGCTGAGAAATATACAGCTGCTCTCAC 888
```

Db 66438 TATGAGCCCACTCCAGATTAGGAACCTATGAGATCCGTTCCCACTCCC 66497
Qy 889 AGATCT 894
Db 66498 AGATAT 66503

RESULT 10
AC104325 224086 bp DNA linear ROD 30-MAY-2003
LOCUS Mus musculus clone rp23-204m3 map 15 strain C57BL/6J, complete
DEFINITION sequence.
AC104325
VERSION AC104325.28 GI:31193955
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Jiang, H., Song, L. and Roe, B.A.
TITLE Mus musculus BAC Clone rp23-204m3
JOURNAL Unpublished
REFERENCE
AUTHORS Jiang, H., Song, L. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2001) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Jiang, H., Song, L. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-2003) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Jiang, H., Song, L. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-2003) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Jiang, H., Song, L. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (30-MAY-2003) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On May 30, 2003 this sequence version replaced gi:29124187.
----- Genome Center
Center: Department of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

FEATURES
source location/Qualifiers
1..224086
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="15"
/clone="rp23-204m3"
/clone_id="RPCT - 23 Female (C57BL/6J) Mouse BAC Library"
BASE COUNT 55592 a 55124 c 56095 g 57275 t
ORIGIN

Query Match 17.5%; Score 156.8; DB 10; Length 224086;
Best Local Similarity 66.4%; Pred. No. 1.7e-28;
Matches 244; Conservative 0; Mismatches 117; Indels 6; Gaps 1;

Qy 529 GTGATCTATGAGCCCACTCCAGATTATGAGATCCAGAGGATATGGAATCCCA 588
Db 162690 GTTCCCTATGAGCCCACTCCAGATTATGAGGCTCACCCTGAGATACGAGATCCCA 162749

Qy 589 CCAGAAGATATGAGACCCCAACCAAGGGGATATGAGAGCCCACTATGAGATATGAGCC 648
Db 162750 TCTGCAAGGATATGAGACCCCACTCCGCGGATATGAGAGCCCACTATGAGATATGAGCC 162809
Qy 649 CCGCCTGAGATATGAGATCCCACTGCGGGATATGAGATCCCACTGCGGGATATGAG 708
Db 162810 CCATCTCCGAGATACGATCTCTGCTCTCGATATGAGAG-----CTGTGAGATATGAGC 162863
Qy 709 GTCCCACTGCGGGATATGAGAGCCCACTGCGGGATATGAGATCCCACTGCGGGATAT 768
Db 162864 TCCCTCTCTCTCTATATATATCTATCCCAATGAGGCTATGAGAGTTCCACTCTGTGATAT 162923
Qy 769 GGTGCCCCCACTGCGGGATATGAGAGCCCACTGCGAGATATGAGAGCCCACTGAG 828
Db 162924 GGAACCTCACTCTGATATGAGATCCCACTCTCGATATGAGAGCCCACTATGAG 162983
Qy 829 AATGAGCCCTTACCCCTGATATGAGATCCCACTCTGATATGAGATATGAGATATGAG 888
Db 162984 TATGAGCCCACTCCAGATATGAGATCCCACTATGAGATATGAGATATGAGATAT 163043
Qy 889 AGATCT 894
Db 163044 AGATAT 163049

RESULT 11
AC107527/c 253149 bp DNA linear HTG 13-MAY-2003
LOCUS Rattus norvegicus clone CH230-92M24, WORKING DRAFT SEQUENCE, 3
DEFINITION unordered pieces.
AC107527
VERSION AC107527.5 GI:30580771
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS 1 (bases 1 to 253149)
Muzny, D., Marle, Metzkler, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, B., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, B.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, Z.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyler, M., Cree, A., D'Souza, L.,
Davila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Drapear, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Baves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Pan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, T., Foster, P.,
Fraser, C. M., Gabriel, A., Ganta, R., Garcia, A., Garner, T., Garza, W.,
Gebregeorgis, B., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Guarinate, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, L., Louie, H., Louie, H., Lozano, R. J., Lu, X., Lu, J.,
Maheshwari, M., Mahindaratne, M., Mahmood, M., Mallory, K., Mangum, A.,
Mangum, B., Mapua, P., McNeill, T. Z., Meenen, B.,
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, B., Moore, S.,
Milošević, A., Miner, G., Minja, B., Montemayor, J., Murphy, M., Naif, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwachukwu, O., Okunolu, G., Olarinmoye, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfamkoeh, C.,

Plapper, F., Polindexter, A., Popovic, D., Primus, R., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, M., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Sma, J., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sores, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trajce, Z., Umami, K., Valera, R., Vera, V., Villaseca, D., Waldron, J., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Wilson, R., Wleczek, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 253149)
Worley, K.C.

Direct Submission
Submitted (23-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 253149)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:23664681.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GKFL
Center clone name: CH230-92M24

----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 242817 bases at least Q40
Consensus quality: 245517 bases at least Q30
Consensus quality: 247516 bases at least Q20
Estimated insert size: 257092; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Ratimmed insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 250781: contig of 250781 bp in length
* 250782 250881: gap of unknown length
* 250882 251966: contig of 1085 bp in length

FEATURES
source * 251967 252066: gap of unknown length
* 252067 253149: contig of 1083 bp in length.
1..253149
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-92M24"

misc_feature
2794..3331
/note="clone boundary
clone_end:77
site:BCORI
end_sequence:BH296484"

misc_feature
2825..3625
/note="clone boundary
clone_end:77
site:BCORI
end_sequence:BH296484"
complement(246782..247500)
/note="clone boundary
clone_end:8p6
site:BCORI
end_sequence:BH296486"

BASE COUNT 64040 a 61494 c 61630 g 62235 t 3750 others
ORIGIN

Query Match 16.9% Score 151.8; DB 2; Length 253149;
Best Local Similarity 65.9%; Pred. No. 3.1e-27;
Matches 238; Conservative 0; Mismatches 117; Indels 6; Gaps 1;

QY 534 CTATGACACCCCAACCAAGATATGACGTCCTCAACGAGGATATGAGTCCACGAA 593
DB 118121 CTATGACACCCCAACGAGGATATGAGTCCACGAGGATATGAGTCCACGAG 118062

QY 594 AGATATGAGACCCCAACGAGGATATGAGTCCACGAGGATATGAGTCCACGAG 653
DB 118061 AGATATGAGACCCCAACGAGGATATGAGTCCACGAGGATATGAGTCCACGAG 118002

QY 654 TGTGGATATGAGTCCACGAGGATATGAGTCCACGAGGATATGAGTCCACGAG 713
DB 118001 TGTGGATATGAGTCCACGAGGATATGAGTCCACGAGGATATGAGTCCACGAG 117948

QY 714 AACTGGGGATATGAGTCCACGAGGATATGAGTCCACGAGGATATGAGTCCACGAG 773
DB 117947 AACTGGGGATATGAGTCCACGAGGATATGAGTCCACGAGGATATGAGTCCACGAG 117888

QY 774 CCCACCTGGGATATGAGTCCACGAGGATATGAGTCCACGAGGATATGAGTCCACGAG 833
DB 117887 CCCACCTGGGATATGAGTCCACGAGGATATGAGTCCACGAGGATATGAGTCCACGAG 117828

QY 834 AGCCCTACCCCTGATATGAGTCCACGAGGATATGAGTCCACGAGGATATGAGTCCACGAG 893
DB 117827 AGCCCTACCCCTGATATGAGTCCACGAGGATATGAGTCCACGAGGATATGAGTCCACGAG 117768

QY 894 T 894
DB 117767 T 117767

RESULT 12
AC132969/c 270171 bp DNA linear HTG 20-NOV-2002
LOCUS AC132969
DEFINITION Rattus norvegicus clone CH230-327L20, WORKING DRAFT SEQUENCE.
ACCESSION AC132969
VERSION AC132969.3 GI:25139203
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FUTUROP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 270171)
Murny, D., Marle, M., Metzger, M., Lee, A., Abramson, S., Adams, C., Alder, J.,

QY	774	CCGACCTGGGGGAATATGAGAGCCCACTGGAGAGATATGAGAGCCCAAGCAGCTGGAAATGA	833
DB	15135	CCGACCTGTGAGATACGGATATCCCACTCTCGATTACGAGAGCCCACTATATGACATATGG	15076
QY	834	AGCCCTACCCCTTCGATATGAAAGCTTCATCTGCTGGAATAATACAGTCCCTTCACAGATC	893
DB	15075	CGCCCACTTCCTCATATACGTAACTACACCTATGGAATCCGGTTCCCACTCCCGGATA	15016
QY	894	T 894	
DB	15015	T 15015	
RESULT 13			
LOCUS	AF499026	793 bp	mRNA linear ROD 23-APR-2002
DEFINITION	Rattus norvegicus Wm-domain binding protein 2 (Wbp2) mRNA, complete cds.		
ACCESSION	AF499026		
VERSION	AF499026.1	GI:20269980	
KEYWORDS			
ORGANISM	Rattus norvegicus (Norway rat)		
REFERENCE	Rattus norvegicus		
AUTHORS	Zannini, M. and Nitsch, R.		
TITLE	Isolation of the cDNA encoding for rat Wbp2		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 793)		
AUTHORS	Zannini, M. and Nitsch, R.		
TITLE	Direct Submision		
JOURNAL	Submitted (04-APR-2002) DBPCW, University of Naples Federico II, via Pansini 5, Naples, NA 80131, Italy		
FEATURES			
SOURCE	location/Qualifiers		
gene	1..793		
CDS	/organism="Rattus norvegicus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10116"		
	1..793		
	/gene="Wbp2"		
	1..789		
	/gene="Wbp2"		
	/codon_start=1		
	/product="Wm-domain binding protein 2"		
	/protein_id="PAM18132.1"		
	/db_xref="GI:20269981"		
	/translation="MALNRKHSREGGVIVNNTESILMSYDHYELTFNDKRVPAFKG TKGQVLTPEPRVIFLSKGDNRSPMPPEYKQDEVQPVFGANFIKGTWKAAG GWSGSASYKLTPTAGAILBERQMLQVAQASRGSEVNPQAYQPYMPSGAYVFP NGMVPKGVYPPPEPRFPGPMDMDGAMGYQPPPEYPPMPBPVVSGSAPFP AARKAAEAAASLYVNPNGNHNVTSGPPEPPPYTPPDKTY"		
BASE COUNT	189 a 259 c 196 g 149 t		
ORIGIN			
Query Match	14.3%; Score 128.4; DB 10; Length 793;		
Best Local Similarity	60.3%; Pred. No. 2, 6e-21;		
Matches 251; Conservative 0; Mismatches 156; Indels 9; Gaps 2;			
QY	1	ATGGCAGTGAACCAAGACCAACCCGAGAGCCGTCGTGGGAGCCCTCATCCCTCTGGCGAA	60
DB	1	ATGGCGCTCAACAAAGATCACTCAGAGGGCGGCGAGTATCGTCAACAACT--GAG	57
QY	61	AGTGTCTTGAAGCAGTGTGAGATGTGAACTCTGCTTCTACAGAAACAGTGAATTC	120
DB	58	AGCATCTTAATGTCTTATGATCATGTGAACTTACATCAACGACATGGAATGACCA	117
QY	121	TATCTCTTTATGSCAAGAAAGAAAGAAAGTGTTCGACTTCATACGGGGTGGCTTC	180
DB	118	GAGGCTTCAAGAGGACCAAGAAAGGACCCGTCTACTCAATCTCAATCCGGGTATCTTT	177
QY	181	GTGACTTCAACTAGTCAATGAGCCCACTTTCTTTATATGATGCCGTTTGAGCTGATG	240

```

Db      178  CTGC-----CAAGGCGAAGAGTCCATGCGGCTCTTCATGATGCGCTTCTTACTATG 231
QY      241  AGTACGTGCACCATTTAAACAACCAATTTTGGCCCCCACTACATTAAAGAACCAATTGAG 300
Db      232  AAGACGTGTGAGGTCAAGACACCGGTGTTTGTGTCCAACTTCATTAAAGGGACAGTGA 291
QY      301  GCAGCTCAGGTGTGTCGTGTGGGAAGAGCAAGCTGTTTAAAGTATCCTTCAGAAAGGA 360
Db      292  GCTAAGAGGAGAGGTGTGTGGGAAGGCTCTGCTCTTAACAAGCTGAACCTTCAACGAGGG 351
QY      361  GTGTCATTCGAATTTGGCCCAACTAGTGGTAAAGCTGCTGCTGCTGCTGCGCAGAGG 416
Db      352  GTGTCCATTGAGTTTGGGACGACAGATCTCCAGGTGGCATCTCAAGCTCCAGAGG 407

RESULT 14
LOCUS   NMU40826                      1802 bp    mRNA    linear    ROD 04-MAR-2003
DEFINITION Mus musculus WW-domain binding protein 2 mRNA, complete cde.
ACCESSION U40826
VERSION   U40826.1 GI:1777578
KEYWORDS Mus musculus (house mouse)
SOURCE   Mus musculus
ORGANISM Mus musculus
REFERENCE 1.
AUTHORS  Bukacynska, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITL     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURN     Sudol, M., Chen, H.I., Bougeret, C., Binbond, A., and Bork, P.
MEDLINE  95369475
PUBMED   7641887
REFERENCE 2. (bases 1 to 1802)
AUTHORS  Chen, H.I. and Sudol, M.
TITL     The WW domain of Yes-associated protein binds a proline-rich ligand
JOURN     that differs from the consensus established for Src homology
MEDLINE  95372370
PUBMED   7644498
REFERENCE 3. (bases 1 to 1802)
AUTHORS  Chen, H.I. and Sudol, M.
TITL     Identification and Characterization of Protein Ligands to the WW
JOURN     domain by Western Ligand Blotting
MEDLINE  95372370
PUBMED   7644498
REFERENCE 4. (bases 1 to 1802)
AUTHORS  Academic Press, Inc., USA (1996)
TITL     TECHNICALS IN PROTEIN CHEMISTRY VII: 3-12;
JOURN     4 (bases 1 to 1802)
MEDLINE  95372370
PUBMED   7644498
REFERENCE 5. (bases 1 to 1802)
AUTHORS  Chen, H.I. and Sudol, M.
TITL     Direct Submission
JOURN     Submitted (17-NOV-1995) Henry I. Chen, Laboratory of Molecular
MEDLINE  95372370
PUBMED   7644498
REFERENCE 6. (bases 1 to 1802)
AUTHORS  Oncology, The Rockefeller University, 1230 York Avenue, New York,
TITL     NY 10021, USA
JOURN     Location/Qualifiers
MEDLINE  95372370
PUBMED   7644498
REFERENCE 7. (bases 1 to 1802)
AUTHORS  1. 1802
TITL     /organism="Mus musculus"
JOURN     /mol_type="mRNA"
MEDLINE  95372370
PUBMED   7644498
REFERENCE 8. (bases 1 to 1802)
AUTHORS  /db_xref="taxon:10090"
TITL     /dev_stage="16 day embryo"
JOURN     40. 825
MEDLINE  95372370
PUBMED   7644498
REFERENCE 9. (bases 1 to 1802)
AUTHORS  /note="WPB-2; putative protein ligand to the WW domain of
TITL     Yes-kinase associated protein YAP"
JOURN     /codon_start=1
MEDLINE  95372370
PUBMED   7644498
REFERENCE 10. (bases 1 to 1802)
AUTHORS  /product="WW-domain binding protein 2"
TITL     /protein_id="AAB40826.1"
JOURN     /db_xref="GI:1777578"
MEDLINE  95372370
PUBMED   7644498
REFERENCE 11. (bases 1 to 1802)
AUTHORS  /translation="MALNKGKSGGGVIVANNSTSLMSYDVELTFNDMKVPEAPFKG
TITL     TKGGVAVLPRVFLSKDQMSFPPVTLSDQCEIKQVPGGANPKIGIVKADAGG
JOURN     GKGSAAYKILFTAGALIRGGQMSLYAGSGDQGRVNGALGVYMSGALVPPPPYNA
MEDLINE  95372370
PUBMED   7644498
REFERENCE 12. (bases 1 to 1802)
AUTHORS  NGRTFPCPGVPIPPPPPPPTTTPPPPCGAGCTGQPPPPPTTTPPPPPVSGSAPATPA
TITL     AEAQAAAPAAASAVITPNSFNPNVNTATVATGPPPPPTTTPPPDKTQ"
JOURN     AEAQAAAPAAASAVITPNSFNPNVNTATVATGPPPPPTTTPPPDKTQ"
MEDLINE  95372370
PUBMED   7644498
REFERENCE 13. (bases 1 to 1802)
AUTHORS  366 A 565 C 437 G 404 T
TITL     BASE COUNT
JOURN     366 A 565 C 437 G 404 T
MEDLINE  95372370
PUBMED   7644498

```

ORIGIN

Query Match 14.3%; Score 128.4; DB 10; Length 1802;
 Best Local Similarity 60.3%; Pred. No. 2.5e-21;
 Matches 251; Conservative 0; Mismatches 156; Indels 9; Gaps 2;

Qy 1 ATGCACTGAGACCAAGACCAACCAAGCGTGTGGGCGCTCATCCCTTGAGGAA 60
 |||||
 Db ATGGCGCTCAAGATCACTGAGAGGGGGCGAGTGTCTCAACAACT---GAG 96
 Qy 61 AGTGTCTTGAAGAGTGTGAGATGTGACCTTGTCTTCTTCAAGAAACAGTGAATCC 120
 |||||
 Db AGCATCTTAATGTCTTATGATGATGAGAGCTTACCTTCAACGACAGAAAGATGCCA 156
 Qy 121 TATCTCTTATGAGCAAAAGAAAGAAAGCTTTTCTCATCTTCCGGGTGCTTC 180
 |||||
 Db GAGGCTTCAAGAGGACCAAGAAAGGACCGTCTTACTTCTGTAACGGGTGCTTTT 216
 Qy 181 GTGACTTCAACTTATGATGACCCCATGCTTTCTTTATGATGCGGTTTGCGCTGATG 240
 |||||
 Db CTGTCTTAAAG---GGAAAGAGCGCATGCACTTCTTCAATGATGCCCTTTTACTGATG 270
 Qy 241 AGTACTGACACATTAAGCAACCAATTTTGGCCCACTACATTAAGAAACCATTCAG 300
 |||||
 Db AAGGACTGTGATTAAGACAGCCGTTGTGTGCGAATTCATTAAGGAAATGTAAG 330
 Qy 301 GCAGCTTCAAGTGTGTGCTGGGAAAGCAAGCTTTTAAAGTATCTTCAAGAAAGCA 360
 |||||
 Db GCTGAGACGAGAGGTGTGCTGGAAAGGCTCCGCTCTTCAAGAGCTTCAAGCAGG 390
 Qy 361 GGTGCTTCAAGATTTGGCCCACTGATGTATTAAGCTGCTGTGCTGCGACAGAG 416
 |||||
 Db GGGCCATTGATTTGGGACAGAGATCTCCAGGTGCACTTCAAGCTTCAAGAG 446

RESULT 15
 BX470149/c 222469 bp DNA linear HTG 07-MAY-2003
 LOCUS BX470149.3 GI:30424228
 DEFINITION Danio rerio clone DKFZ-242K7, *** SEQUENCING IN PROGRESS ***, 50
 unoriented pieces.

ACCESSION BX470149
 VERSION BX470149.3 GI:30424228
 KEYWORDS HTG; HTGS PHASE1.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 222469)

REFERENCE
 AUTHORS Burton, J.
 TITLES Direct Submission
 JOURNAL Submitted (06-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

COMMENT
 On May 7, 2003 this sequence version replaced gi:30387077.

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfish-help@sanger.ac.uk
 ----- Project Information
 Center project name: zK242K7
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 197958 bases at least Q40
 Consensus quality: 206715 bases at least Q30
 Consensus quality: 212351 bases at least Q20
 Insert size: 217569; sum-of-coverage
 Insert size: 165837; 6.2% error; agarose-fp
 Quality coverage: 2.48x in Q20 bases; sum-of-coverage
 coverage: 3.95x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 50 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 6656: contig of 6656 bp in length
 6657 gap of 100 bp
 6757 10255: contig of 3499 bp in length
 10256 gap of 100 bp
 10355 13223: contig of 2868 bp in length
 13224 gap of 100 bp
 13324 19935: contig of 6612 bp in length
 19936 gap of 100 bp
 20036 22597: contig of 2562 bp in length
 22598 gap of 100 bp
 22599 22697: contig of 5872 bp in length
 22698 gap of 100 bp
 28570 37965: contig of 9296 bp in length
 37966 gap of 100 bp
 38066 41140: contig of 3075 bp in length
 41141 gap of 100 bp
 41241 45108: contig of 3868 bp in length
 45109 gap of 100 bp
 45209 47730: contig of 2522 bp in length
 47731 gap of 100 bp
 47831 51884: contig of 4054 bp in length
 51885 gap of 100 bp
 51985 55857: contig of 3873 bp in length
 55858 gap of 100 bp
 55958 60498: contig of 4541 bp in length
 60499 gap of 100 bp
 60599 63641: contig of 3043 bp in length
 63642 gap of 100 bp
 63742 72674: contig of 8933 bp in length
 72675 gap of 100 bp
 72775 75602: contig of 2828 bp in length
 75603 gap of 100 bp
 75703 78062: contig of 2360 bp in length
 78063 gap of 100 bp
 78163 84485: contig of 6323 bp in length
 84486 gap of 100 bp
 84586 87553: contig of 2966 bp in length
 87554 gap of 100 bp
 87654 92487: contig of 4834 bp in length
 92488 gap of 100 bp
 92588 96104: contig of 3517 bp in length
 96105 gap of 100 bp
 96205 101479: contig of 5275 bp in length
 101480 gap of 100 bp
 101580 104979: contig of 3400 bp in length
 104980 gap of 100 bp
 105080 108003: contig of 2924 bp in length
 108004 gap of 100 bp
 108104 113151: contig of 5048 bp in length
 113152 gap of 100 bp
 113153 116585: contig of 3334 bp in length
 116586 gap of 100 bp
 116587 119346: contig of 2661 bp in length
 119347 gap of 100 bp
 119447 123125: contig of 3679 bp in length
 123126 gap of 100 bp
 123226 131265: contig of 8041 bp in length
 131267 gap of 100 bp
 131367 136193: contig of 4827 bp in length
 136194 gap of 100 bp
 136293 139482: contig of 3189 bp in length
 139483 gap of 100 bp
 139582 141875: contig of 2293 bp in length
 141876 gap of 100 bp
 141975 144122: contig of 2147 bp in length

*	144223:	144223:	gap of 100 bp
*	144223:	145966:	contig of 276 bp in length
*	144569:	147068:	gap of 100 bp in length
*	147069:	149399:	contig of 231 bp in length
*	149400:	149499:	gap of 100 bp
*	149500:	153963:	contig of 4464 bp in length
*	153964:	154063:	gap of 100 bp
*	154064:	156579:	contig of 2516 bp in length
*	156580:	156679:	gap of 100 bp
*	156680:	158928:	contig of 2249 bp in length
*	158929:	159028:	gap of 100 bp
*	159029:	164542:	contig of 5514 bp in length
*	164543:	164642:	gap of 100 bp
*	164643:	172055:	contig of 7412 bp in length
*	172055:	172155:	gap of 100 bp in length
*	172155:	175210:	contig of 3056 bp in length
*	175211:	175310:	gap of 100 bp
*	175311:	179379:	contig of 4069 bp in length
*	179380:	179479:	gap of 100 bp
*	179480:	182698:	contig of 3219 bp in length
*	182699:	182798:	gap of 100 bp
*	182799:	186779:	contig of 3991 bp in length
*	186780:	186879:	gap of 100 bp
*	186880:	195628:	contig of 8747 bp in length
*	195629:	195726:	gap of 100 bp
*	195727:	199396:	contig of 3670 bp in length
*	199397:	199496:	gap of 100 bp
*	199497:	206412:	contig of 6916 bp in length
*	206413:	206512:	gap of 100 bp
*	206513:	209619:	contig of 3107 bp in length
*	209620:	209719:	gap of 100 bp
*	219084:	219088:	contig of 9364 bp in length
*	219184:	222465:	gap of 100 bp in length
*	222466:	222469:	contig of 3266 bp in length

```
misc_feature      /organism="Danio rerio"  
                  /mol_type="genomic DNA"  
                  /db_xref="taxon:7955"  
                  /clone="DKX-242K7"  
                  /clone_id="Dan10key"  
1. 6656  
/note=assembly_fragment:01359  
fragment chain:1"  
misc_feature      6757..10255  
/note=assembly_fragment:01385  
fragment chain:1"  
misc_feature      10356..13223  
/note=assembly_fragment:00552  
fragment chain:1"  
misc_feature      13324..19935  
/note=assembly_fragment:00237  
fragment chain:1"  
misc_feature      20036..32597  
/note=assembly_fragment:00949  
fragment chain:1"  
misc_feature      32698..28169  
/note=assembly_fragment:00615  
fragment chain:1"  
misc_feature      28670..37965  
/note=assembly_fragment:01306  
fragment chain:2"  
misc_feature      38066..41140  
/note=assembly_fragment:00487  
fragment chain:2"  
misc_feature      41241..45108  
/note=assembly_fragment:00232  
fragment chain:2"  
misc_feature      45309..47730  
/note=assembly_fragment:00935  
fragment chain:2"  
misc_feature      47831..51884  
/note=assembly_fragment:01415
```

```

misc_feature      fragment chain:2"
                    51985..55857
                    /note="assembly_fragment:00878
                    fragment chain:3"
                    55858..60498
                    /note="assembly_fragment:01208
                    fragment chain:3"
                    60509..63641
                    /note="assembly_fragment:01266
                    fragment chain:3"
                    63742..73674
                    /note="assembly_fragment:01329
                    fragment chain:3"
                    73775..75602
                    /note="assembly_fragment:01183
                    fragment chain:4"
                    75703..78062
                    /note="assembly_fragment:01179

```

Query Match	13.7%	Score 122.8;	DB 2;	Length 222469;
Best Local Similarity	88.7%	Pred. No. 5.8e-20;		
Matches 133; Conservative	0;	Mismatches 17;	Indels 0;	Gaps 0;

Qy	171	GGTGGTCGCGAGCTTACGACTAGTCAAAAGACCCAGCTTCCTTATGATGCGGT	230
Db	190751	GGTATCTTCGTGACTTACACTCAGTCGACGACCCCAAGTTTCTTTCATATAGCAAT	1906922
Qy	231	TGGCTGATGAGTGACTGCACCATTTGAACCAATTTTGGCCCCCACTACATTAAGG	290
Db	190691	TGATCTGATGAGTAACTGACACATTTGAACAAACCCGCTTTTGCCCCCACTACATTAAGG	1906322
Qy	291	AACCATTCAGGACGCTCCAGGATGAGCGG	320
Db	190631	AACGCTTCAGGACGCTCCAGATGATGAGTG	190602

Search completed: December 16, 2003, 15:10:10
Job time : 3343.84 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 05:54:30 / Search time 239.646 Seconds

(Without alignments)
10115.338 Million cell updates/sec

Title: US-09-864-291-4_COPY_36_933

Perfect score: 898
Sequence: 1 atgscagtgtagaccagagcca.....tgcctctcacagatctatga 898

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Genseq_19Jun03:*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	898	100.0	1413	24	AAS20601	DNA encoding bovin
2	374.4	41.7	1001	22	AAS20602	Human encoding human
3	262.6	29.2	467	22	AAU01229	Human reproductive
4	262.6	29.2	467	23	ABL96688	Human testicular a
5	220	24.5	7099	22	AAU04882	Human reproductive
6	220	24.5	7099	23	ABL97776	Human testicular a
7	220	24.5	220895	24	ABR84798	Human CDNA differe
8	216.8	24.1	436	22	ABA67850	Human foetal liver

9	216.8	24.1	436	22	AAK42003	Human bone marrow
10	216.8	24.1	436	22	AAI48070	Probe #16756 used
11	216.8	24.1	436	24	ABR16034	Human genome-deriv
12	195.4	21.8	321	23	AAU01355	Human reproductive
13	195.4	21.8	321	23	ABL96688	Human testicular a
14	122	13.6	1885	24	ABR70430	Human bone remodel
15	122	13.6	1885	21	AAE22363	Human secreted pro
16	122	13.6	1915	25	ABR73614	Secreted protein-e
17	122	13.6	1915	25	ABR16841	Human secreted pro
18	122	13.6	1915	25	ABR67208	Human secreted pro
19	121.2	13.5	894	23	AAS70582	Human secreted pro
20	114.4	12.7	409	25	ABX41819	DNA encoding novel
21	103.8	11.6	454	21	AAU04803	Bovine EST associa
22	92.4	10.3	291	23	AAU04883	Human secreted pro
23	92.4	10.3	291	23	ABR97777	Human reproductive
24	86	9.6	471	22	ABR55246	Human testicular a
25	86	9.6	471	22	AAK28957	Human foetal liver
26	86	9.6	471	22	AAI34906	Human bone marrow
27	86	9.6	471	22	AAI34906	Probe #3592 used t
28	73.6	8.2	420	24	ABR76335	Human genome-deriv
29	70	7.8	136	21	AAK22819	Bacillus lichenifo
30	68.8	7.7	591	22	ABR49404	Human secreted pro
31	68.8	7.7	591	22	ABR47315	Human breast cell
32	68.8	7.7	591	22	ABR44407	Human foetal liver
33	68.8	7.7	591	22	AAK15750	Probe #12873 for g
34	68.8	7.7	591	22	AAK41489	Human bone marrow
35	68.8	7.7	591	22	AAI22235	Human bone marrow
36	68.8	7.7	591	22	AAI47531	Probe #12168 for g
37	68.8	7.7	591	22	AAI07935	Probe #7926 used t
38	68.8	7.7	591	23	ABR41079	Human liver singl
39	68.8	7.7	591	24	ABR15494	Human genome-deriv
40	68.8	7.7	1959	22	ABR44250	Human breast cell
41	68.8	7.7	1959	22	ABR45699	Human foetal liver
42	68.8	7.7	1959	22	ABR44480	Probe #2946 for ge
43	68.8	7.7	1959	22	AAK02984	Human brain expres
44	68.8	7.7	1959	22	AAK28431	Human bone marrow
45	68.8	7.7	1959	22	AAI12997	Probe #2930 for ge

ALIGNMENTS

RESULT 1	AAI20601	standard, cDNA, 1413 BP.
AC	AAI20601;	
XX	09-APR-2002	(first entry)
XX	DNA encoding bovine perinuclear theca 32 (PT32).	
DE	Testicular WW domain binding protein; hMWBP; perinuclear theca 32;	
XX	PT32; contraceptive; fertility; oocyte activation; vaccine;	
KW	globozoospermy; spermatogenesis; spermatozoa; tyrosine kinase; c-Yes;	
KM	immunoreceptor; bovine; gene; ss.	
XX	Bos sp.	
OS		
XX		
PH	Key	Location/Qualifiers
FT	primer_bind	30..50
FT		/tag= a
FT		/note= "Primer binding site for cDNA isolation. The
FT		sequence differs from that of the forward primer
FT		given in AAS20603"
FT	CDS	36..977
FT		/tag= b
FT		/product= "PT32"
FT		/note= "Perinuclear theca 32"
FT	protein_bind	978..1001
FT		/tag= c
FT		/note= "Primer binding site for cDNA isolation. The
FT		sequence differs from that of the reverse primer

XX PD 29-NOV-2001.
 XX PF 25-MAY-2001, 2001MO-CA00738.
 XX PR 25-MAY-2000, 2000CA-2307128.
 XX PR 25-MAY-2000, 2000US-206979P.
 XX PA (TOOH) UNIV QUEBENS KINGSTON.
 XX PA (UYOR-) UNIV OREGON HEALTH SCI.
 XX PI Olo R, Sutovsky P;
 XX DR WPI: 2002-097644/13.
 XX DR P-PSDB; AAU74610.
 PT Isolated perinuclear theca 32 polypeptide that interacts with activated
 PT tyrosine kinase c-Yes, for enhancing fertility, creating/diagnosing
 PT diminished fertility and abnormal spermiogenesis and for providing
 PT contraception -
 XX
 XX PS Claim 62; Fig 4B; 103pp; English.
 CC The invention describes an isolated perinuclear theca 32 (PT32)
 CC polypeptide (I) which interacts with tyrosine kinase c-Yes. (I) is
 CC useful for: enhancing fertility in a mammal; treating globozoospermy, by
 CC expressing (I) in spermatozoa; inhibiting fertilisation, by introducing
 CC (I) or its antigenic fragment into a mammal to elicit an immune
 CC response; enhancing the ability of round spermatids to activate oocytes;
 CC treating or diagnosing diminished fertility and abnormal spermiogenesis;
 CC in providing contraception; identifying contraceptive and
 CC fertility-enhancing agents. The polynucleotide is useful for producing
 CC (I) by recombinant techniques, as vaccine, as diagnostic reagents, and
 CC for chromosome identification. An antibody against (I) is useful in
 CC immunological assays, in immunoneutralization methods, to identify cells
 CC expressing (I), and to purify (I) by affinity chromatography. A
 CC transgenic animal is useful as an animal model for studying human
 CC fertility and reproductive biology, and for screening compounds to
 CC identify components of oocyte activation. The use of (I) prevents the
 CC entry of components which are detrimental to embryonic development into
 CC the oocyte during oocyte activation with crude sperm extract and avoids
 CC the propagation of viruses such as HIV (human immunodeficiency virus) and
 CC SIV (simian immunodeficiency virus) carried in the sperm. This sequence
 CC encodes the human testicular MW domain binding protein (hMBP), described
 CC in the method of the invention.
 XX
 XX SQ Sequence 1001 BP; 261 A; 254 C; 248 G; 238 T; 0 other;
 Query Match 41.7%; Score 374.4; DB 24; Length 1001;
 Best Local Similarity 75.3%; Pred. No. 5,1e-101;
 Matches 497; Conservative 0; Mismatches 151; Indels 12; Gaps 2;
 QY 223 ATGCGCTTGGCTGATGAGTGAAGTGCACCATTTGAACAACATTTTGGCCCCCACTAC 282
 DB 1 ATGCCATTGATCTGATGAGTGAAGTGCACCATTTGGAACAACATTTTGGCAAACTTC 60
 QY 283 ATTAAGGAACCATTTAGGAGAGTGCAGAGTGTGGCTGGGAGAGAGCAAGCTTTTAAAG 342
 DB 61 ATTAAGGAACCATTTAGGAGAGTGCAGAGTGTGGCTGGGAGAGAGCAAGCTTTTAA 120
 QY 343 TTATCCCTTCAGAGAAAGAGTGTGCATGAAATTTGCCCACTGATGTAAGGTGCTCT 402
 DB 121 TTAGTCTTCAGAAATGGAATGCAATGCAATTTGCGCAATGTAAGGTGCTCT 180
 QY 403 GGTGCTGCGAGAGAAATTCACCTTGAAGTGAATTAATTAATTAATTAATTAATTAATTA 462
 DB 181 GGTGCTGCGAGAGAAATTCACCTTGAAGTGAATTAATTAATTAATTAATTAATTAATTA 240
 QY 463 TACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 522
 DB 241 TATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 292
 QY 523 CCAATGTGATCTATGAGACCCCAACCAAGATATATAGTCAACCAAGGAGAAATATGGA 582

DB 293 ----TTATGTCTATAGGAGGCCCACTGCAAGATATAGAGCCCACTCCCGAATACGA 348
 QY 583 ACTCCACGAGAGGATATAGAGGCCCAACCAAGGAGATATAGAGGCCCACTTATAGGATAT 642
 DB 349 GCTCCACCTGCAAGATATAGAGGCCCAACCTGATAGAAATATAGAGGCCCGCTTGGGATAC 408
 QY 643 GGAGCCCGCTGTGAGATATAGATCCCACTGGAGGAGATATAGAGTCCCACTGGAGGGA 702
 DB 409 AAGGCTCACTGTGAGATATAGAGGCCCACTCTTGTGATACGAGGCCCACTTCAGAGA 468
 QY 703 TATGAGTCCCACTGGAGGATATAGAGGCCCACTGGAGGATATAGAGTCCCACTGGAG 762
 DB 469 TATGAGGCCCACTCTGAGATATAGAGGCCCACTCTGATATAGAACCCCACTCTC 528
 QY 763 GGATATAGTCCCACTGGAGGATATAGAGGCCCACTCTGAGATATAGAGGCCCACTCTC 822
 DB 529 GGATATAGAGGCCCACTCTGAGATATAGAGGCCCACTCTGAGAAATATAGAGGCCCGCT 588
 QY 823 GCTGGAATGAGAGCCCTACCCCTGATATAGAGTCCATCTGCTGAAATACAGCTGCC 882
 DB 589 GCGGATACAGAGCCCTCACTGCTGATCAGAGGCCAGGCTCAGAAATCTACAGAGCC 648
 RESULT 3
 AAL01229
 ID AAL01229 standard; cDNA; 467 BP.
 XX
 XX AAL01229;
 AC
 AC 21-NOV-2001 (first entry)
 DT
 XX
 DE Human reproductive system related antigen cDNA SEQ ID NO: 1230.
 XX
 XX Human, reproductive system related antigen; reproductive system disorder;
 XX cancer; gene therapy, ss.
 XX
 XX Homo sapiens.
 XX
 XX MO200155320-A2.
 PD
 PD 02-AUG-2001.
 XX
 XX 17-JAN-2001; 2001MO-US01339.
 XX
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216413.
 PR 07-JUL-2000; 2000US-0216847.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225265.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227109.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231245.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 25-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI, 2001-46570/50.
 DR P-PSDB; AAM95259.
 DR
 XX
 XX Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 XX
 PS Claim 1; SEQ ID NO 1230; 1297bp + Sequence listing; English.
 CC
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a coding sequence of the
 CC invention.
 XX
 SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
 Query Match 29.2%; Score 262.6; DB 22; Length 467;
 Best Local Similarity 76.4%; Pred. No. 7.2e-68;
 Matches 365; Conservative 0; Mismatches 100; Indels 13; Gaps 3;
 QY 351 CAGAAAGAGGTCATCATGATTTGCCCACTGATGTAAAGCTGCTCTGCTGC 410
 Db 2 CAGAAATGAGGCAATGAAATTTGCCAGTGAATGTAAAGCTGCTCTGCTGC 61
 QY 411 CAGAGGATTCGACTTGGAAGTGAATTAATGAGTGTGACACTTCAGACTGTACATTAAT 470
 Db 62 CCGAGATTTCCACTTGAATCTTAATGACTGTGCTGCTGAGCTTATGGAATTTATGTAAT 121
 QY 471 TACTGTCCAGGGGCTGCAAGTGTGCTCTCAGACAGACCTTGTCCAGCATATCAATGT 530
 Db 122 TACTGGGGAAGGAAT---ATGTGCACTCAGACAGATGCTTGTTCAG-----TAT 169
 QY 531 GATCTATGACCCCAACACCAAGATATACAGTCAACAGGGGAATATGAACTCAC 590

DB 170 TGTCTATGAGCCCGCCACTGACAGATATGAGCCCGCCCGGATAGAGAGCCCGCCACC 229
QY 591 AGAAGATATGAGAGCCCAACAGAGGATATGAGAGCCCGCCACTTATGAGATATGAGAGCCCG 650
DB 230 TGCAGATATGAGAGCCCGCCCGCTAGAAATGAAAGCCCGCTGTGATACAGAGCCCTC 289
QY 651 GCTGTGATATGAGAGCCCGCCCGGATATGAGAGCCCGCCCGGATATGAGAGCT 710
DB 290 ACTGTGATATGAGAGCCCGCCCGCTTGTATGAGAGCCCGCCCGCTGACAGATATGAGAGC 349
QY 711 CCCACTGTGAGATATGAGAGCCCGCCACTGTGAGATATGAGAGCTCCGCTGTGAGATATG 769
DB 350 CCCACTGTAGATATGAGAGCCCGCCCGCTTGTATGAGATATGAGAGCCCGCTTGTGATATG 409
QY 770 GTGCCCCACTGTGAGATATGAGAGCCCGCCCGCTGACAGATATGAGAGCCCGCCACTGTG 827
DB 410 GAGCCCCACTGTGTGATATGAGAGCCCGCCCGCTGACAGAAATGAAAGCCCGCTTGTGAG 467

RESULT 4

ABL96688

ID ABL96688 standard; cDNA; 467 BP.

XX ABL96688;

XX 21-JUN-2002 (first entry)

XX Human testicular antigen encoding cDNA SEQ ID NO: 356.

XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;

XX reproductive system disorder; urinary system disorder; gene therapy;

XX gastrovascular disorder; respiratory disorder; neurological disorder;

XX gastrointestinal disease; infection; cytostatic; gene; ss.

XX Homo sapiens.

XX WO200155317-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01329.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216447.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 14-JUL-2000; 2000US-0217496.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483332/52.
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
XX useful for preventing, diagnosing and/or treating testicular cancer -
XX
XX
XX Claim 1; SEQ ID NO 356; 768bp; English.
XX
XX The present invention provides the protein and coding sequences of 973
XX human testicular antigens, and fragments of their genomic sequences. The
XX sequences can be used in the treatment of cardiovascular, urinary system,
XX reproductive system, immune, respiratory, neurological and
XX gastrointestinal disorders, infections, and particularly cancer,
XX especially testicular cancers. The present sequence is a cDNA of the
XX invention.
XX
XX
XX Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
SQ
Query Match 29 2%; Score 262.6; DB 23; Length 467;
Best Local Similarity 76.4%; Pred. No. 7.2e-68; Indels 13; Gaps 3;
Matches 365; Conservative 0; Mismatches 100;
QY 351 CAGAAAGAGGTGCATTCGATTTGCCCACTGATGTAAGCTGCTCTGCTGCTGC 410
DB 2 CAGAAATGAGGTGCATTCGATTTGCCCACTGATGTAAGCTGCTCTGCTGCTGC 61
QY 411 CAGAGAAATTCACCTTGGAAGTGAATTACTGTTGCACTTCAGACATGTACATAT 470
DB 62 CCGAGATTTCACTAGAACCTTAAGACGTTCAGCTCTATGAGGAATTTATGTAAT 121
QY 471 TACTGTCCAGGGGTGAGTGTGCTCCACAGACACTGTTCAGACATTCGAATGT 530
DB 122 TACTGGGAGGGAAT---ATGTGCACTTCACAGATGCTTGTGAG-----TTAT 169
QY 531 GATCTATGACCCCAACCAACAGATATACAGTCCCAAGGGAATATGAACTCCACC 590
|||||

DB 170 TGTCTATGAGCCCACTGACAGATATGAGCCCACTCCCGATACGAGCCCAACC 229
QY 591 AGAAGATATGAGCCCAACAGGGGATATGAGCCCACTATGAGGATATGAGCCCC 650
DB 230 TGCAGATATGAGCCCAACCTTAGAATGAGGCCCTGTGAGATACAGAGCTTC 289
QY 651 GCCTGTGAGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATGAGT 710
DB 290 ACCGTGCGATATGAGCCCACTTGTGATAGGAGCCCACTGCGAGATATGAGC 349
QY 711 CCCACTGAGGATATGAGCCCACTGAGGATATGAGTCCCACTGAGGATATG 769
DB 350 CCCACTGATGATATGAGCCCACTTGTGATATGAGACCCCACTCTGAGATATG 409
QY 770 GTGCCCACTGAGGATATGAGCCCACTGACAGATATGAGCCCAACGCTG 827
DB 410 GAGCCCACTTGTGATATGAGCCCACTGACAGATATGAGGCGGCTTTCGG 467
RESULT 5
ID AAL04882 standard; DNA; 7099 BP.
XX
XX AAL04882;
XX
XX 21-NOV-2001 (first entry)
XX
XX Human reproductive system related antigen DNA SEQ ID NO: 7570.
XX
XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX
XX Homo sapiens.
XX
XX MO200155320-A2.
XX
XX
XX 02-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US01339.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0186874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.

Qy 819 ACCAGCTGGAATGAAGCCTACCCCTGCATATGAGCTCCATCTGCGAATATACAGC 878
Db 7018 GCCTCGGGATACAGAGCCTCACCTGCGAATCAGAGCCAGAGCCTCAGAAATCTACAGC 7077
Qy 879 TGCC 882
Db 7078 AGCC 7081

RESULT 6
ABL97776
ID ABL97776 standard; DNA; 7099 BP.
XX ABL97776;
AC
XX 21-JUN-2002 (first entry)
DT
XX
DE Human testicular antigen encoding DNA fragment SEQ ID NO: 2428.
XX
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
XX reproductive system disorder; urinary system disorder; gene therapy;
XX cardiovascular disorder; respiratory disorder; neurological disorder;
XX gastrointestinal disease; infection; cytostatic; gene; de.
OS Homo sapiens.
XX
XX WO200155317-A2.
PD
XX 02-AUG-2001.
XX
PP 17-JAN-2001; 2001MO-US01329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249220.
 PR 17-NOV-2000; 2000US-0249221.
 PR 17-NOV-2000; 2000US-0249222.
 PR 17-NOV-2000; 2000US-0249223.
 PR 17-NOV-2000; 2000US-0249224.
 PR 17-NOV-2000; 2000US-0249225.
 PR 17-NOV-2000; 2000US-0249226.
 PR 17-NOV-2000; 2000US-0249227.
 PR 17-NOV-2000; 2000US-0249228.
 PR 17-NOV-2000; 2000US-0249229.
 PR 17-NOV-2000; 2000US-0249230.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0256097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483332/52.
 XX
 PT Nucleic acids encoding 973 human testicular antigen polypeptides,
 XX useful for preventing, diagnosing and/or treating testicular cancer -
 PS Disclosure; SEQ ID NO 2428; 766bp; English.
 XX
 CC The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer.
 CC especially testicular cancers. The present sequence is a DNA encoding a
 CC protein fragment of the invention.
 XX
 SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
 Query Match 24.5%; Score 220; DB 23; Length 7099;
 Best Local Similarity 75.3%; Pred. No. 1.4e-54;
 Matches 274; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
 QY 519 ATATCCATTGTAATCTATGAGCCGCCCAACAGAGATATACATCCCAAGGGGAAATA 578
 DB 6718 ATTCCTGATATATGCTATGAGCCGCCCACTGAGATATGAGCCCACTCCCGAATA 6777
 QY 579 TGAACCTCCACAGAGATATGAGCCCAACAGAGATATGAGCCCACTATAGG 638
 DB 6778 CGAGACCCCACTGAGATATGAGCCCAACAGAGATATGAGCCCACTCTGTGG 6837
 QY 639 ATATGAGCCCTGCTGTGAGATATGAGTCCCACTTGGGGATATGAGTCCCACTGG 698
 DB 6838 ATACGAGACCTCACTGTGAGATATGAGCCCACTCTGTGATATGAGCCCACTGCG 6897
 QY 699 GGGATATGAGTCCCACTGAGATATGAGCCCACTTGGGGATATGAGTCCCACTGG 758
 DB 6898 AGGATATGAGCCCACTGAGATATGAGCCCACTCTGTGATATGAGCCCACTGCG 6957
 QY 759 TGGGGATATGAGTCCCACTGAGATATGAGCCCACTGAGATATGAGCCCACTGG 818
 DB 6958 TCTGAGATATGAGCCCACTCTGTGAGATATGAGCCCACTGAGATATGAGCCCACTGG 7017
 QY 819 ACCAGCTGAAATGAAGCCCTTACCCCTGATATGAGCTCATCTGCTGAGAAATACAGC 878
 DB 7018 GCTGCGGGATATGAGAGCTTCACTGTGAGATATGAGAGCGAGCTTCAAGATATACAGC 7077

QY 879 TGCC 882
 DB 7078 AGCC 7081
 RESULT 7
 ABK84798
 ID ABK84798 standard, cDNA; 220895 BP.
 XX
 AC ABK84798;
 XX
 DT 14-AUG-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #1369.
 XX
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; AIDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 XX
 OS Homo sapiens.
 XX
 PN WO200228999-A2.
 XX
 XX 11-APR-2002.
 XX
 PF 03-OCT-2001; 2001WO-US30821.
 XX
 PR 03-OCT-2000; 2000US-237189P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX WPI; 2002-435328/46.
 XX
 DR Detecting granulocyte activation by detecting differential expression
 XX of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -
 XX
 PS Claim 1; SEQ ID NO 1369; 114bp; English.
 XX
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression (M2) Gs is indicative of GCA.
 CC Also included are modulating (M3) Gs by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC or allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating Gs; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal

reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's disease, ulcerative colitis, periodontal disease, also bacterial infection, viral infection, CC paratitic infection, protozoal infection, fungal infection and MS is useful for treating one of the above conditions. The present CC sequence represents a gene differentially expressed in granulocytes. CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic CC format directly from WIPO at CC ftp.wipo.int/pub/published_pct_sequences.

Sequence 220895 BP; 53760 A; 54808 C; 55831 G; 56496 T; 0 other;

Query Match 24.5%; Score 220; DB 24; Length 220895;
Best Local Similarity 75.3%; Pred. No. 8.4e-54;
Matches 274; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

519 ATATCCAAATGATGATCTATGAGACCCCAACAGATATACATCCACAGGGGATA 578
169998 ATTCCAGTTATGTCTATGAGCCCACTGAGATATGAGCCCACTCCCGGATA 170057
579 TGGAACTCCACAGAGATATGAGCCCAACAGGGGATATGAGCCCACTTATGG 638
170058 CGGAGCCCACTGAGAGATATGAGCCCAACCGTGAAGAAATGAGCCCGCTTGGG 170117
639 ATATGAGACCCCGCTGTGGATATGAGTCCCACTGGGGATATGAGTCCCACTGG 698
170118 ATACAGAGCTTCACTGTGATATGAGCCCACTTGTGATATGAGCCCACTTGC 170177
699 GGGATATGAGTCCCACTGAGGGATATGAGCCCACTGGGGATATGAGTCCCACT 758
170178 AGGATATGAGAGCCCACTCTAGATATGAGCCCACTTGTGATATGAGCCCACT 170237
759 TGGGGATATGAGTCCCACTGAGGGATATGAGCCCACTGAGAGATATGAGCCCACT 818
170238 TCTGGATATGAGAGCCCACTCTGAGATATGAGCCCACTGAGAAATGAGCCCACT 170297
819 ACCGCTGAGAAATGAGCCCACTGAGATATGAGTCCCACTGAGAAATGAGC 878
170298 GCTGCGGAGATACAGAGCTTCACTGTGATATGAGAGCCCACTGAGAAATGAGC 170357
879 TGCC 882
170358 AGCC 170361

RESULT 8

ABA67850 ID ABA67850 standard; DNA; 436 BP.

AC ABA67850;

XX 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #16155.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLR-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI, 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -

XX Claim 4; SEQ ID NO 16155; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC fetal liver. The present sequence is a single exon nucleic acid

CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;

Query Match 24.1%; Score 216.8; DB 22; Length 436;
Best Local Similarity 75.6%; Pred. No. 3.1e-54;
Matches 269; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

527 TTGTGATCTATGAGACCCCAACAGATATACATCCACAGGGGATATGAGAACTC 586
1 TTATTTCTATGAGAGCCCACTGAGAGATATGAGCCCACTCCCGATATGAGAGCC 60
587 CACAGAGATATGAGAGCCCAACAGAGGATATGAGAGCCCACTATGAGATATGAG 646
61 CACCTGAGATATGAGAGCCCAACAGAGGATATGAGAGCCCACTATGAGATATGAG 120
647 CCCGCTGTGAGATATGAGTCCCACTGAGGGATATGAGTCCCACTGAGGGATATG 706
121 CTTCACTGTGAGATATGAGAGCCCACTTGTGATATGAGAGCCCACTGAGAGATATG 180
707 GAGTCCCACTGAGGGATATGAGAGCCCACTGAGGGATATGAGAGTCCCACTGAGGGAT 766
181 GAGCCCACTCTAGAGATATGAGAGCCCACTTGTGATATGAGAGCCCACTCTGAGAT 240
767 ATGATCCCACTGAGGGGATATGAGAGCCCACTGAGAGATATGAGAGCCCACTGAG 826
241 ATGAGAGCCCACTCTGAGATATGAGAGCCCACTGAGAGATATGAGAGCCCACTGAG 300
827 GAAATGAGAGCCCACTGAGATATGAGAGTCCATCTGAGAAATGAGAGTCCG 882
301 GATACAGAGCTTCACTGAGATATGAGAGCCCACTGAGAGATATGAGAGTCCG 356

RESULT 9

AAK42003 ID AAK42003 standard; DNA; 436 BP.

XX AAK42003;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 16560.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

XX microarray; cancer; leukemia; lymphoma; myeloma; ss.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow.
 XX
 PS Example 4; SEQ ID NO: 16560; 658bp + sequence listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 24.1%; Score 216.8; DB 22; Length 436;
 Best Local Similarity 75.6%; Pred. No. 3.1e-54;
 Matches 269; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 QY 527 TTGTATCTATGAGACCCCAACGAGATATACAGTCCACGAGGGAATATGAACTC 586
 DB 1 TTATGTCTATGAGACCCCAACGAGATATGAGACCCCACTCCGATACGAGCCC 60
 QY 587 CACGAGAAGATATGAGACCCCAACGAGGGAATATGAGACCCCACTATGAGATATGAG 646
 DB 61 CACTGCAAGATATGAGACCCCAACGAGGGAATATGAGACCCCACTGATACAGAG 120
 QY 647 CCCCCTGTGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATG 706
 DB 121 CCTCACTGTGGAATATGAGACCCCACTCTTGATATGAGACCCCACTGCAAGATATG 180
 QY 707 GAGTCCCACTGAGGGAATATGAGACCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 766
 DB 181 GAGCCCACTCTAGGATATGAGACCCCACTCTTGATATGAGACCCCACTCTCGAT 240
 QY 767 ATGTGACCCCACTGAGGGAATATGAGACCCCACTGAGGGAATATGAGACCCCACTG 826
 DB 241 ATGTGACCCCACTCTGAGATATGAGACCCCACTGCAAGAAATGAGACCCCACTGAG 300
 QY 827 GAAATGAAGCCCTACCCCTGATATGAGTCCATCTGAGAAATATAGAGTACC 882
 DB 301 GATACAGAGCTCACTGCTGATATGAGACCCCACTGAGAAATATAGAGAGCC 356
 RESULT 10
 AAI48070
 ID AAI48070 standard; DNA; 436 BP.
 XX
 AC AAI48070;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #16756 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 XX genetic disorder; 88.
 OS Homo sapiens.
 XX

PN W0200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PP 30-JAN-2001; 2001WO-US000663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-48897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 16756; 654bp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENPs).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 24.1%; Score 216.8; DB 22; Length 436;
 Best Local Similarity 75.6%; Pred. No. 3.1e-54;
 Matches 269; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 QY 527 TTGTATCTATGAGACCCCAACGAGATATACAGTCCACGAGGGAATATGAACTC 586
 DB 1 TTATGTCTATGAGACCCCACTGCAAGATATGAGACCCCACTCCGATACGAGCCC 60
 QY 587 CACGAGAAGATATGAGACCCCAACGAGGGAATATGAGACCCCACTATGAGATATGAG 646
 DB 61 CACTGCAAGATATGAGACCCCAACGAGGGAATATGAGACCCCACTGATACAGAG 120
 QY 647 CCCCCTGTGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATG 706
 DB 121 CCTCACTGTGGAATATGAGACCCCACTCTTGATATGAGACCCCACTGCAAGATATG 180
 QY 707 GAGTCCCACTGAGGGAATATGAGACCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 766
 DB 181 GAGCCCACTCTAGGATATGAGACCCCACTCTTGATATGAGACCCCACTCTCGAT 240
 QY 767 ATGTGACCCCACTGAGGGAATATGAGACCCCACTGAGGGAATATGAGACCCCACTG 826
 DB 241 ATGTGACCCCACTCTGAGATATGAGACCCCACTGCAAGAAATGAGACCCCACTGAG 300
 QY 827 GAAATGAAGCCCTACCCCTGATATGAGTCCATCTGAGAAATATAGAGTACC 882
 DB 301 GATACAGAGCTCACTGCTGATATGAGACCCCACTGAGAAATATAGAGAGCC 356
 RESULT 11
 ABS16034
 ID ABS16034 standard; DNA; 436 BP.
 XX
 AC ABS16034;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human genome-derived single exon probe ORF from lung SEQ ID No 16025.
 XX
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 XX

KM chronic obstructive pulmonary disease; interstitial lung disease;
 KM familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KM tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KM Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemostidrosis;
 KM pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome;
 KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KM primary ciliary dyskinesia; pulmonary hypertension;
 KM hyaline membrane disease; open reading frame; ORF.
 XX
 OS Homo sapiens.
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2002-114183/15.
 DR
 XX Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 PT
 PS Claim 4; SEQ ID No 16025; 634bp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridize at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridization of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridization to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences mentioned in the specification, or encoded by the
 CC probe/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemostidrosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Kargener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 24.1%; Score 216.8; DB 24; Length 436;
 Best Local Similarity 75.6%; Pred. No. 3.1e-54;
 Matches 269; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 QY 527 TTGTGATCTATGAGCCCCCAACAGATATACAGTCCAAACGAGGGAATATGAACTC 586
 DB 1 TTATGCTATGAGCCCCCACTGACGATATGAGACCCCACTCCCGATACGAGCCC 60
 QY 587 CACCAAGAGATATGAGCCCAACGAGGGAATATGAGACCCCACTATAGGATATGAG 646
 DB 61 CACCTCAGATATGAGCCCAACCGTAGAAATATGAGCCCTGAGGATATACAGAG 120
 QY 647 CCCGACCTGAGATATGAGTCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATG 706
 DB 121 CTCACCTGATATGAGCCCACTCTTGAATACGAGCCCACTGACGATATG 180
 QY 707 GAGTCCCACTGAGGGAATATGAGCCCACTGAGGGAATATGAGTCCCACTGAGGGAAT 766
 DB 181 GAGCCCACTGAGATATGAGCCCACTCTGGAATATGAGACCCCACTCTGGAT 240
 QY 767 ATGTGCCCCCACTGAGGGAATATGAGCCCACTGACGATATGAGCCCACTGAG 826
 DB 241 ATGAGGCCCACTCTGAGATATGAGCCCACTGACGAAATATGAGGCCCTGAGG 300
 QY 827 GAATGAGCCCTACCCCTGATATGAGCTCATCTGCTGGAATATACAGTACC 882
 DB 301 GATACAGGCTTACCTGCTGATCAGAGCCAGGCTCAGGAATCTACAGCAGCC 356
 RESULT 12
 ID AAL01355 standard; cDNA; 321 BP.
 XX
 AC AAL01355;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DB Human reproductive system related antigen cDNA SEQ ID NO: 1356.
 XX
 KM Human; reproductive system related antigen; reproductive system disorder;
 KM cancer; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200155320-A2.
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01339.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.

PR	14-AUG-2000	2000US-0225157
PR	14-AUG-2000	2000US-0225158
PR	14-AUG-2000	2000US-0225159
PR	18-AUG-2000	2000US-0226671
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226682
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227009
PR	30-AUG-2000	2000US-0229824
PR	01-SEP-2000	2000US-0229817
PR	01-SEP-2000	2000US-0229943
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229545
PR	05-SEP-2000	2000US-0229593
PR	05-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231442
PR	08-SEP-2000	2000US-0231443
PR	08-SEP-2000	2000US-0231444
PR	08-SEP-2000	2000US-0231445

ID ABL96808 standard; cDNA; 321 BP.

DT 21-JUN-2002 (first entry)

DE Human testicular antigen encoding cDNA SEQ ID NO: 476.

Human, teicular antigen, testee; cancer; metastasis; immune disorder
reproductive system disorder; urinary system disorder; gene therapy;
cardiovascular disorder; respiratory disorder; neurological disorder;
gastrointestinal disease; infection; cytostatic; gene; ss.

OS Homo sapiens

PN W0200155317-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01329.

PR 31-JAN-2000; 2000US-0179065.

PR 24-FEB-2000; 2000US-0184664.

PR 16-MAR-2000; 2000US-0189874.

PR 18-APR-2000; 2000US-0198123.

PR 07-JUN-2000; 2000US-0209467.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225447.

PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246612.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Rosen CA, Barash SC, Ruben SW,
XX WFI, 2001-463232/52.
XX
DR Nucleic acids encoding 973 human testicular antigen polypeptides,
PT useful for preventing, diagnosing and/or treating testicular cancer -
XX
XX
PS Claim 1; SEQ ID NO 476; 766bp; English.
XX
XX The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a cDNA of the
CC invention.
XX
SQ Sequence 321 BP; 83 A; 75 C; 70 G; 89 T; 4 other;

Query Match	Similarity	21.8%	Score 195.4	DB 23	Length 321
Best Local	Similarity	77.4%	Pred. No. 6,3e-48		
Matches	246	Conservative	2	Matches 69	Indels 1
Qy	29	GGCGTCGAGGAGCCCTCATCCCTCTGCGGAAAGTCTTGAAGCAGTGAAGATGTG	88		
Db	1	GGCGCGCGAGACCCCTCATCCCTTAACGGTGAAGTCTTGAAGCGGCTCCGAATGG	60		
Qy	89	ACCTCTGCTTCTTAACGAAACAGTGAATCTATCTTTTAATGGCAAAAGAA	148		
Db	61	ACCTCTCTTCCACAGCGATCAGAAAGCTCAAAATGCTTTAGGTGAAGAACAGAA	120		
Qy	149	CGTTGTTTTCACCTTATACGGGGGTCTTGTGACTTCAACCTTAATGATGACCCA	208		
Db	121	CATTGTTTTCACCTTATACGGGTGATTTTATTAATTCATGCTTCAACAGTATGCCA	180		

Qy	209	TGCTTCTTTATATGATGCGTTTGCGCTGATGATGATGATCTGACACATTTAAACAACCAATTT	268
Db	161	TGTATCTTTTATATGATGCGATTTATCTGATATACGAACCTGACCTGTGTAAACAACCAATTT	240
Qy	269	TTGCCCCCAACTACATTAAGGAGACATTCAAGCAGCTCCAGGTGTGTGCTGGG- AAGA	327
Db	241	TTGCTGCACAACTTCATTAAGGAGAACTATTCAAGCAGCTTCATATGAGGCGCTGGGAAGA	300
Qy	328	CAAGCTGTTTTAAAGTTA	345
Db	301	CAAGCTACTTTAAATA	318

RESULT 14

ID ABS70430 standard; cDNA; 1885 BP.

AC ABS70430;

DT 27-NOV-2002 (first entry)

DB Human bone remodelling gene #87.

KM Bone remodelling; osteoporosis; human; gene; ss.

OS Homo sapiens.

PN US6426186-B1.

PD 30-JUL-2002.

PR 18-JAN-2000; 2000US-0484970.

PR 18-JAN-2000; 2000US-0484970.

PA (INCY-) INCYTE GENOMICS INC.

PI Jones KA, Volkmuth W, Walker MG;

DR WPI; 2002-673014/72.

PT A combination of polynucleotides which are co-expressed with genes
PT known to be involved in bone remodeling and osteoporosis are useful in
PT an array for the diagnosis of bone remodeling and osteoporosis
PT associated disorders -

PS Claim 1; Column 247-250; 206pp; English.

CC The invention relates to a combination comprising a number of
CC substantially purified and isolated polynucleotides which are
CC co-expressed with genes known to be involved in bone remodelling and
CC osteoporosis. The invention is used to diagnose disorders associated
CC with bone remodelling or osteoporosis. AB570344-AB570512 represent
CC human bone remodelling genes of the invention.

SQ Sequence 1885 BP; 384 A; 610 C; 476 G; 415 T; 0 other;

Query Match	13.6%	Score 122;	DB 24;	Length 1885;
-------------	-------	------------	--------	--------------

Matches 247; Conservative 0; Mismatches 160; Indels 9; Gaps 2;

1 ATGGCAGTGAACGAGGCCACACGAGAGCCGTCGTGGGGCCCTCATCCCTCTGGCGAA 60

Db 61 ATGGCGCTCACAGATCACTCGAGGGCGCGGAGTATCTCAATAACAC---CGAG 117

61 AGTGTCTGAAGCAGTGTGAGATGTGACCTCTGCTTCTCTACAGAAACCAGTGGAATCC 120

Db 118 AGCATCTAATGTCCTATGATCAGTGAATCATTCAATGACATGAGAAGTCCCA 177

121 TATCTCTTAA TGCAAGAAGCAAGCTTGTCTCTCACTTCATATACCGGTCCTTC 180

Db 178 GAAGCCTTCAAAGGACCAAGAAAGCACTGTCTACCTTACCCCTTACCGGTCATCTT 237

QY 181 GTGACTTCACACTTAAGTACGACCCGAGCTTTCTTTATGATGCGCTTGAGC 240
 DB 238 CTGTC-----CAAGGCGAAGATGCCATGACGCTTCAATGATGACATTTATTCAG 231
 QY 241 AGTACTGACCACTTGAACAACCAATTTTGTCCCACTACATTAAGAACATTCAG 300
 DB 292 AAGACTGTGATGATCAAGACGCGGATTTGTGTGCAAACTACATCAAGGAACAGTGAAG 351
 QY 301 GCAGCTCCAGTGTGTGCTGGGAAGACAACTGTTTAAATTCCTTCAGAAAGGA 360
 DB 352 GCGAAGCGGAGAGTGTGCTGGGAAGGCTCTGCTCCACAAATTTGACTTTCAGCGGAGG 411
 QY 361 GTGCGCATGGAATTTGCGCACTGATGTAAGTGAAGCTGCTGTGCTGCGCAGAG 416
 DB 412 GCGCGCATTTGATGTTGCGAAGCGGATGCTCAAGTGTGACATTCAGGCTTCAGAGG 467

RESULT 15

AAF22363
 ID AAF22363 standard; cDNA; 1915 BP.

AC AAF22363;

DT 26-MAR-2001 (first entry)

DE Human secreted protein gene 48 SEQ ID NO:58.

XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 XX autoimmune; antiproliferative; cytostatic; cardiac; vasotropic;
 XX cerebroprotective; neuroprotective; antibacterial; virucide;
 XX fungicide; ophthalmological; vulnery; gene therapy; neoplasm;
 XX autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 XX cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 XX cerebral ischaemia; angiogenesis; nervous system disorder; infection;
 XX Alzheimer's disease; ocular disorder; corneal infection; wound healing;
 XX skin aging; food additive; preservative; ss.

OS Homo sapiens.

PN WO200061748-A1.

PD 19-OCT-2000.

PF 06-APR-2000; 2000MO-US08982.

XX 09-APR-1999; 99US-0128696.

PR 14-JAN-2000; 2000US-0176069.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-638566/61.

DR P-PSDB; AAB63096.

XX New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -

XX Claim 1; Page 429-430; 480pp; English.

CC AAF22316 to AAF22363 encode the human secreted proteins given in AAB63049
 CC to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins
 CC and polypeptides homologous to them. Human secreted proteins have
 CC activities based on the tissues and cells the genes are expressed in.
 CC Examples of activities include: immunosuppressive; antiarthritic;
 CC antineumatic; antiproliferative; cytostatic; cardiac; vasotropic;
 CC cerebroprotective; neuroprotective; antibacterial; virucide;
 CC fungicide; ophthalmological; and vulnery. The polynucleotides and
 CC proteins can be used to prevent, treat or ameliorate a medical condition
 CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
 CC sheep. They are also used in diagnosing a pathological condition or

CC susceptibility to a pathological condition. Disorders which are diagnosed
 CC or treated include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angioneitis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
 CC ocular disorders e.g. corneal infection. The polypeptides can also be
 CC used to aid wound healing and epithelial cell proliferation, to prevent
 CC skin aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis. The polypeptides can also be used as a food additive or
 CC preservative to increase or decrease storage capabilities, fat content,
 CC lipid, protein, carbohydrate, vitamins, minerals, cofactors and other
 CC nutritional components. AAF22307 to AAF22315 and AAB63048 represent
 CC sequences used in the exemplification of the present invention.

SQ Sequence 1915 BP; 401 A; 613 C; 481 G; 413 T; 7 other;

Query Match 13.6%; Score 122; DB 21; Length 1915;
 Best Local Similarity 59.4%; Pred. No. 1.2e-25;
 Matches 247; Conservative 0; Mismatches 160; Indels 9; Gaps 2;

QY 1 ATGGCAGTGAACCAAGCCACACCGAGCGCTGTGGGCGCTCATCCCTTGGCGAA 60
 DB 59 ATGGCGCTCAACAAGATCACTCGAGGGCGGAGTATGTCATTAACAC--CGAG 115
 QY 61 AGTGTCTTGAAGCAGTGTGAGATGTGACCTCTGTCTTCTACAGAAACAGTGAATCC 120
 DB 116 AGATCTTAAATGCTCTATATATACCGTGAACCTCAATTAATGACATGAAGACGCGCA 175
 QY 121 TATCTCTTAAATGGCAAGAAGAAAGCGTTGTTCTCACTTACACGGGTGCTTC 180
 DB 176 GAAGCTTTAAAGGACCAAGAAAGGACGTGTACTTACCTTACCGGTATCTTT 235
 QY 181 GTGACTTCACACTTAAGTCAATGACCCCATGCTTTCTTTATGATGCGGTTGGCTGATG 240
 DB 236 CTGTC-----CAAGGCGAAGATGTCATGCAAGTCTTCATATGTCATTTATTCAG 289
 QY 241 AGTACTGACCACTTGAACAACCAATTTTGTCCCACTACATTAAGAACATTCAG 300
 DB 290 AAGACTGTGATGATCAAGACGCGGATTTGTGCAAACTACATCAAGGAACAGTGAAG 349
 QY 301 GCAGCTCCAGTGTGTGCTGGGAAGACAACTGTTTAAATTCCTTCAGAAAGGA 360
 DB 350 GCGAAGCGGAGAGTGTGCTGGGAAGGCTGTCTCTTCAAGTTGACTTTCAGCGAGG 409
 QY 361 GTGCGCATGGAATTTGCGCACTGATGTAAGTGAAGCTGCTGTGCTGCGCAGAGG 416
 DB 410 GCGCGCATTTGATGTTGCGAAGCGGATGCTCAAGTGTGACATTCAGGCTTCAGAGG 465

Search completed: December 16, 2003, 11:00:47
 Job time : 241.646 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: December 16, 2003, 09:55:15 ; Search time 63.2646 Seconds
(without alignments)
6265.145 Million cell updates/sec

Title: US-09-864-291-4_COPY_36_933
Perfect score: 898
Sequence: 1 atgcgcagtcagccagagccca.....tgccctcacagatcatga 898

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents NA:
2: /cgm2_6/prodata/2/ina/5A_COMB.seq:*
3: /cgm2_6/prodata/2/ina/5B_COMB.seq:*
4: /cgm2_6/prodata/2/ina/6A_COMB.seq:*
5: /cgm2_6/prodata/2/ina/6B_COMB.seq:*
6: /cgm2_6/prodata/2/ina/PCTUS_COMB.seq:*
7: /cgm2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122	13.6	1885	4	US-09-84-970B-87 Sequence 87, Appl
2	59.8	6.7	3489	2	US-08-728-323A-1 Sequence 1, Appl
3	59.8	6.7	3489	4	US-09-298-568-1 Sequence 1, Appl
4	59.8	6.7	3489	4	US-09-410-399-1 Sequence 1, Appl
5	59.8	6.7	32207	2	US-08-770-379-20 Sequence 20, Appl
6	59.8	6.7	32207	3	US-08-757-669A-20 Sequence 20, Appl
7	59.8	6.7	32207	4	US-09-230-371A-20 Sequence 20, Appl
8	58.4	6.5	981	2	US-08-841-349-15 Sequence 15, Appl
9	54.8	6.1	3833	2	US-08-917-320-18 Sequence 18, Appl
10	54.8	6.1	3833	5	PCT-US95-04611A-18 Sequence 18, Appl
11	54.8	6.1	5931	4	US-08-783-774-1 Sequence 1, Appl
12	54.8	6.1	5931	4	US-09-556-706B-1 Sequence 1, Appl
13	53.4	5.9	243	1	US-08-182-175A-56 Sequence 56, Appl
14	53.4	5.9	243	1	US-08-474-633A-74 Sequence 74, Appl
15	53.4	5.9	243	1	US-08-823-771-74 Sequence 74, Appl
16	53.4	5.9	243	5	PCT-US92-06412-56 Sequence 56, Appl
17	53	5.9	2144	3	US-08-834-306-15 Sequence 15, Appl
18	53	5.9	2144	4	US-08-993-674A-15 Sequence 15, Appl
19	53	5.9	2144	4	US-09-256-976-15 Sequence 15, Appl
20	52.4	5.8	2793	1	US-08-209-747-1 Sequence 1, Appl
21	52.4	5.8	2793	1	US-08-458-298-1 Sequence 1, Appl
22	51.2	5.7	3794	4	US-09-192-434-1 Sequence 1, Appl
23	49.8	5.5	1235	2	US-08-557-309B-53 Sequence 53, Appl
24	49.6	5.5	187	1	US-08-182-175A-90 Sequence 90, Appl
25	49.6	5.5	187	1	US-08-474-633A-78 Sequence 78, Appl
26	49.6	5.5	187	4	US-08-823-771-78 Sequence 78, Appl
27	49.6	5.5	187	5	PCT-US92-06412-90 Sequence 90, Appl

28	49.4	5.5	8310	3	US-08-870-126-11 Sequence 11, Appl
29	49.4	5.5	8310	4	US-09-445-247-11 Sequence 11, Appl
30	49.4	5.5	14985	1	US-08-652-972A-6 Sequence 6, Appl
31	49.4	5.5	14985	5	PCT-US96-06231A-6 Sequence 6, Appl
32	49.2	5.5	2188	1	US-07-865-662P-10 Sequence 10, Appl
33	49.2	5.5	2188	1	US-08-374-219B-10 Sequence 10, Appl
34	48.8	5.4	5661	4	US-08-938-105-2 Sequence 2, Appl
35	48.6	5.4	16442	3	US-08-781-891-208 Sequence 208, App
36	48.6	5.4	16442	4	US-08-166-166-208 Sequence 208, App
37	48.6	5.4	51259	3	US-08-781-891-209 Sequence 209, App
38	48.6	5.4	51259	4	US-09-618-166-209 Sequence 209, App
39	47.4	5.3	471	4	US-09-370-838-278 Sequence 278, App
40	47.4	5.3	977	6	US-08-938-105-2 Sequence 2, Appl
41	46.2	5.1	456	2	US-08-557-309B-16 Sequence 16, Appl
42	46.2	5.1	456	2	US-08-834-306-16 Sequence 16, Appl
43	46.2	5.1	456	3	US-08-993-674A-16 Sequence 16, Appl
44	46.2	5.1	456	4	US-09-256-976-16 Sequence 16, Appl
45	45.8	5.1	977	1	US-08-017-522A-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-484-970B-87
Sequence 87, Application US/09484970B
Patent No. 6426186
GENERAL INFORMATION:
APPLICANT: Jones, Karen A.
APPLICANT: Volkmuth, Wayne
APPLICANT: Walker, Michael G.
TITLE OF INVENTION: BONE REMODELING GENES
FILE REFERENCE: PB-0014 US
CURRENT APPLICATION NUMBER: US/09/484,970B
CURRENT FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 172
SOFTWARE: PERL Program
SEQ ID NO 87
LENGTH: 1885
TYPR: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6426186 232968.10C81
US-09-484-970B-87

Query Match	13.6%	Score 122	DB 4	Length 1885
Best Local Similarity	59.4%	Pred. No. 7.9e-27		
Matches 247	Conservative	0	Mismatches 160	Indels 9
			Gaps	2
QY	1	ATGGCAGTGAACCAAGCCACACCGAGCCCTGCGGGCCCTATCCCTCTGCGGAA	60	
DB	61	ATGGCGCTCAACAAAGATCTACGAGGGCGGAGATGATGTCATTAACAC--CGAG	117	
QY	61	ATGTCCTTAAGCAGTGTGAGATGTGACCTCTGCTTCTACAGAAACAGTGAATCC	120	
DB	118	AGCATCTTAATGTCTATATATCATCTGAACCTCACTTAAGCATGAAGACGTGCA	177	
QY	121	TATCTCTTAATGACCAAGAAAGAAAGAGTGTCTTCACTTACCCGAGTGTCTTC	180	
DB	178	GAGACCTTCAAGAGGACCAAGAAAGGACGTGCTACCTTACCCCTTACGGGGTCACTTT	237	
QY	181	GTGACTTCACTTATGATCAATGACCCCATGCTTTCTTTATATATGCGCTTGGCTGATG	240	
DB	238	CTGTC-----CAAGGCAAGATGCAAGAGTCTTCTATATATGCAATTTATCTCAG	291	
QY	241	ATGACTGCAACATTGAACCAATTTTGGCCCACTAATTAAGAAACATTGAG	300	
DB	292	AAAGACTTGAATCAAGACGCCGATTTGTGCAACTATCAAGGAACAGTGAG	351	
QY	301	CGAGCTCCAGTGTGCTGAGAAAGCAAGCTGTTTAACTTATCTTCAAGAAAGGA	360	
DB	352	GCGAAGCCGAGGATGTGCTGGAAGGCTGTGCTTCAAGATTGACTTCAAGGAGG	411	

QY 361 GCGCCATCGAATTGCGCCCACTGATGTAAGCTGCTGCTGCGCAGAG 416
DB 412 GCGCCATGAGTTCGACAGCGAGTCTCCAGGTGCGATCTCAAGCTCCAGAG 467

RESULT 2
US-08-728-323A-1

/ Sequence 1, Application US/08728323A
/ Patent No. 5948676
/ GENERAL INFORMATION:
/ APPLICANT: Chang, Yuan
/ APPLICANT: Bohenzky, Roy A.
/ APPLICANT: Russo, James J.
/ APPLICANT: Edelman, Isidore S.
/ APPLICANT: Moore, Patrick S.
/ TITLE OF INVENTION: Immediate Early Protein From Kaposi's
/ TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
/ TITLE OF INVENTION: Encoding Same And Uses Thereof
/ NUMBER OF SEQUENCES: 21
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSEE: Cooper & Dunham LLP
/ STREET: 1185 Avenue of the Americas
/ CITY: New York
/ STATE: New York
/ COUNTRY: U.S.A.
/ ZIP: 10036

/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/728,323A
/ FILING DATE:
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: White, John P.
/ REGISTRATION NUMBER: 28,678
/ REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 212-278-0400
/ TELEFAX: 212-391-0525
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3489 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..3489
/ US-08-728-323A-1

Query Match 6.7%; Score 59.8; DB 2; Length 3489;
Best Local Similarity 46.9%; Pred. No. 5.9e-08;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 479 CAGGGGCTGAGTGTGCTCTCA CAGACCTGTGCGAGCATTCACATTTGTATGATG 538
DB 2185 CAGCAGCAGCAGATGACAGCAGCAGATGACAGCAGCAGCAGATGACAGCAG 2244
QY 539 GACCCCAACCAAGATATACAGTCCAA CCGAGGGAATATGAATCTCCACAGAGAT 598
DB 2245 CAGCAGATGAAACAGAGCAGCAGAGAGAGCAGAGCAGAGAGCAGAGCAGAG 2304
QY 599 ATGAGCCCAACAGAGGGAATATGAGCCCA CTTATGGATATGAGCCCTGCTTGG 658
DB 2305 TTACAGGAGCAGAGCAGAGGTTAGAGATCAGAGCAGAGAGTTAGAGAGCAGAG 2364
QY 659 GATATGAGTCCCATCTGGGGATATGAGATCCCATCTGGGGATATGAGATCCCATCTG 718

DB 2365 GAGTTAGAGAGCAGAGCAGAGATTAGAGACAGAGCAGAGATTAGAGAGCAGAG 2424
QY 719 GGGGATATGAGCCCACTGAGGGAATATGAGATCCCATCTGGGGATATGAGTGGCCAC 778
DB 2425 CAGAGTTAGAGAGCAGAGCAGAGATTAGAGACAGAGCAGAGATTAGAGAGCAG 2484
QY 779 CTGGGGATATGAGCCCACTGAGGGAATATGAGAGCCCA CAGAGCTGGAATGAAAGCCC 838
DB 2485 GAGCAGAGATTAGAGAGCAGAGCAGAGATTAGAGAGCAGAGAGTTAGAGAGCAGAG 2544
QY 839 TACCCCTGATATGAGCTCCATCTGCTGGAATATCAG 877
DB 2545 CAGAGGTGGAAGAGCAAGAGCAGAGCAGAGAGAGAGCAG 2583

RESULT 3
US-09-298-568-1

/ Sequence 1, Application US/09298568
/ Patent No. 6323792
/ GENERAL INFORMATION:
/ APPLICANT: Kieff, Elliott D.
/ APPLICANT: Ballesca, Mary E.
/ APPLICANT: Kaye, Kenneth M.
/ TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
/ TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
/ FILE REFERENCE: 16412-10001R
/ CURRENT APPLICATION NUMBER: US/09/298,568
/ CURRENT FILING DATE: 1999-04-21
/ EARLIER APPLICATION NUMBER: US 60/109,422
/ EARLIER FILING DATE: 1998-11-19
/ NUMBER OF SEQ ID NOS: 3
/ SOFTWARE: Patent Ver. 2.0
/ SEQ ID NO 1
/ LENGTH: 3489
/ TYPE: DNA
/ ORGANISM: Kaposi's sarcoma-associated herpesvirus
/ US-09-298-568-1

Query Match 6.7%; Score 59.8; DB 4; Length 3489;
Best Local Similarity 46.9%; Pred. No. 5.9e-08;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 479 CAGGGGCTGAGTGTGCTCTCA CAGACCTGTGCGAGCATTCACATTTGTATGATG 538
DB 2185 CAGCAGCAGCAGATGACAGCAGCAGATGACAGCAGCAGCAGATGACAGCAG 2244
QY 539 GACCCCAACCAAGATATACAGTCCAA CCGAGGGAATATGAATCTCCACAGAGAT 598
DB 2245 CAGCAGATGAAACAGAGCAGCAGAGAGAGCAGAGCAGAGCAGAGCAGAGCAGAG 2304
QY 599 ATGAGCCCAACAGAGGGAATATGAGCCCA CTTATGGATATGAGCCCTGCTTGG 658
DB 2305 TTACAGGAGCAGAGCAGAGGTTAGAGATCAGAGCAGAGAGTTAGAGAGCAGAG 2364
QY 659 GATATGAGTCCCATCTGGGGATATGAGATCCCATCTGGGGATATGAGATCCCATCTG 718
DB 2365 GAGTTAGAGAGCAGAGCAGAGATTAGAGAGCAGAGCAGAGATTAGAGAGCAGAG 2424
QY 719 GGGGATATGAGCCCACTGAGGGAATATGAGATCCCATCTGGGGATATGAGTGGCCAC 778
DB 2425 CAGAGTTAGAGAGCAGAGCAGAGATTAGAGAGCAGAGCAGAGATTAGAGAGCAG 2484
QY 779 CTGGGGATATGAGCCCACTGAGGGAATATGAGAGCCCA CAGAGCTGGAATGAAAGCCC 838
DB 2485 GAGCAGAGATTAGAGAGCAGAGCAGAGATTAGAGAGCAGAGAGTTAGAGAGCAGAG 2544
QY 839 TACCCCTGATATGAGCTCCATCTGCTGGAATATCAG 877
DB 2545 CAGAGGTGGAAGAGCAAGAGCAGAGCAGAGAGAGAGCAG 2583

RESULT 4
US-09-410-399-1

Sequence 1, Application US/09410399
Patent No. 6482587
GENERAL INFORMATION:
APPLICANT: Robertson, Eric S.
APPLICANT: Cotter, Murray A.
TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
FILE REFERENCE: UM-03778
CURRENT APPLICATION NUMBER: US/09/410,399
CURRENT FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 1
LENGTH: 3489
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-1

Query Match 6.7%; Score 59.8; DB 4; Length 3489;
Best Local Similarity 46.9%; Pred. No. 5.9e-08;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 479 CAGGGGCTGCAAGTGTCTCTCAAGACACCTTGTCCAGCATTCATTTGTATCTATG 538
DB 2185 CAGCAGCAGCAGAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAG 2244
QY 539 GACCCCAACCAAGGATATACAGTCCAAACAGGGGATATGAACTCCACCAAGAGAT 598
DB 2245 CAGCAGATGAAACAGAGCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2304
QY 599 ATGAGCCCAACAGGGGATATGAGAGCCCACTATGGATATGAGAGCCCGCTGTG 658
DB 2305 TTAGAGAGCAGAGCAGAGAGATTAAGATCAAGAGCAGAGAGATTAAGAGAGAGAGAG 2364
QY 659 GATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 718
DB 2365 GAGTTAGAGAGCAGAGCAGAGAGATTAAGAGAGAGAGAGAGAGATTAAGAGAGAGAG 2424
QY 719 GGGGATATGAGAGCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 778
DB 2425 CAGAGATTAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2484
QY 779 CTGGGGATATGAGAGCCCACTGAGGATATGAGAGCCCACTGAGGATATGAGAGCC 838
DB 2485 GAGCAGAGATTAGAGAGCAGAGAGAGAGAGAGAGATTAGAGAGAGAGAGAGAGAGAG 2544
QY 839 TACCCCTGCATATGAGCTCATCTGCTGGAATAACAG 877
DB 2545 CAGAGAGTGAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2583

RESULT 5
US-08-770-379-20/c
Sequence 20, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELEPHONE: (212) 391-0525
TELEFAX: (212) 278-0400
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match 6.7%; Score 59.8; DB 2; Length 32207;
Best Local Similarity 46.9%; Pred. No. 1.7e-07;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 479 CAGGGGCTGCAAGTGTCTCTCTCAAGACACCTTGTCCAGCATTCATTTGTATCTATG 538
DB 19812 CAGCAGCAGCAGAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAG 19753
QY 539 GACCCCAACCAAGGATATACAGTCCAAACAGGGGATATGAACTCCACCAAGAGAT 598
DB 19752 CAGCAGATGAAACAGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 19693
QY 599 ATGAGCCCAACAGGGGATATGAGAGCCCACTATGGATATGAGAGCCCGCTGTG 658
DB 19692 TTAGAGAGCAGAGCAGAGAGATTAAGATCAAGAGCAGAGAGATTAAGAGAGAGAGAG 19633
QY 659 GATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 718
DB 19632 GAGTTAGAGAGCAGAGCAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 19573
QY 719 GGGGATATGAGAGCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 778
DB 19572 CAGAGATTAGAGAGCAGAGAGAGAGAGAGAGATTAGAGAGAGAGAGAGAGAGAGAG 19513
QY 779 CTGGGGATATGAGAGCCCACTGAGGATATGAGAGCCCACTGAGGATATGAGAGCC 838
DB 19512 GAGCAGAGATTAGAGAGCAGAGAGAGAGAGAGAGATTAGAGAGAGAGAGAGAGAGAG 19453
QY 839 TACCCCTGCATATGAGCTCATCTGCTGGAATAACAG 877
DB 19452 CAGAGAGTGAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 19414

RESULT 6
US-08-757-669A-20/c
Sequence 20, Application US/08757669A
Patent No. 6183751
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York

STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/757,669A
 FILING DATE:
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 45185-F
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0525
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 32207 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-757-669A-20

Query Match 6.7%; Score 59.8; DB 3; Length 32207;
 Best Local Similarity 46.9%; Pred. No. 1.7e-07;
 Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

479 CAGGGGCTGAGTGTCTCTCTCAAGACCTGTGCTGCAATATTCATTTGATCTATG 538
 DB 19812 CAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAG 19753
 QY 539 GACCCCAACCAAGATATACAGTCCACCAAGGGAATATGAACTCCACCAAGAGAT 598
 DB 19752 CAGCAGATGAAACAGAGCAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAG 19693
 QY 599 ATGAGCCCAACCAAGGGAATATGAGCCCACTTATGAGATATGAGCCCGCTGTGG 658
 DB 19692 TTAGAGAGCAGAGCAGAGGATTAGAGATCAGAGCAGAGGTTAGAGAGCAGAGCAG 19633
 QY 659 GATATGAGATCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTG 718
 DB 19632 GAGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAGAG 19573
 QY 719 GGGGATATGAGAGCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACT 778
 DB 19572 CAGGAGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAG 19513
 QY 779 CTGGGGGATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGAGCC 838
 DB 19512 GAGCAGAGATTAAGAGAGCAGAGCAGAGGTTAGAGAGCAGAGGTTAGAGAGCAGAG 19453
 QY 839 TACCCCTGATATGAGAGCTCCATCTGCTGGAATATCAG 877
 DB 19452 CAGAGGTGAGAGCAGAGCAGAGCAGAGCAGAGGAGAGCAG 19414

RESULT 7
 US-09-230-371A-20/c

Sequence 20, Application US/09230371A
 Patent No. 6348586
 GENERAL INFORMATION:
 APPLICANT: Chang, Yuan
 APPLICANT: Bohenzky, Roy A
 APPLICANT: Russo, James J
 APPLICANT: Beelman, Isidore S
 APPLICANT: Moore, Patrick S
 TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
 TITLE OF INVENTION: US95 THERMOF

FILE REFERENCE: 45185-G-PCT-US
 CURRENT APPLICATION NUMBER: US/09/230,371A
 CURRENT FILING DATE: 1999-11-17
 PRIOR APPLICATION NUMBER: PCT/US97/13346
 PRIOR FILING DATE: 1997-07-22
 NUMBER OF SEQ ID NOS: 30
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 20
 LENGTH: 32207
 TYPE: DNA
 ORGANISM: Kaposi's sarcoma-associated herpesvirus
 US-09-230-371A-20

Query Match 6.7%; Score 59.8; DB 4; Length 32207;
 Best Local Similarity 46.9%; Pred. No. 1.7e-07;
 Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

479 CAGGGGCTGAGTGTCTCTCTCAAGACCTGTGCTGCAATATTCATTTGATCTATG 538
 DB 19812 CAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAG 19753
 QY 539 GACCCCAACCAAGATATACAGTCCACCAAGGGAATATGAACTCCACCAAGAGAT 598
 DB 19752 CAGCAGATGAAACAGAGCAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAG 19693
 QY 599 ATGAGCCCAACCAAGGGAATATGAGCCCACTTATGAGATATGAGCCCGCTGTGG 658
 DB 19692 TTAGAGAGCAGAGCAGAGGATTAGAGATCAGAGCAGAGGTTAGAGAGCAGAGCAG 19633
 QY 659 GATATGAGATCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACTG 718
 DB 19632 GAGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAGAG 19573
 QY 719 GGGGATATGAGAGCCCACTGAGGGAATATGAGTCCCACTGAGGGAATATGAGTCCCACT 778
 DB 19572 CAGGAGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAGAGCAGAGGTTAGAGAGCAG 19513
 QY 779 CTGGGGGATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGAGCC 838
 DB 19512 GAGCAGAGATTAAGAGAGCAGAGCAGAGGTTAGAGAGCAGAGGTTAGAGAGCAGAG 19453
 QY 839 TACCCCTGATATGAGAGCTCCATCTGCTGGAATATCAG 877
 DB 19452 CAGAGGTGAGAGCAGAGCAGAGCAGAGCAGAGGAGAGCAG 19414

RESULT 8

US-08-841-349-15/c
 Sequence 15, Application US/08841349B
 Patent No. 5955594
 GENERAL INFORMATION:
 APPLICANT: MISHRA, LOPA
 TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
 FILE REFERENCE: X/PO470050
 CURRENT APPLICATION NUMBER: US/08/841,349B
 CURRENT FILING DATE: 1997-04-30
 NUMBER OF SEQ ID NOS: 18
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 15
 LENGTH: 981
 TYPE: DNA
 ORGANISM: Mus musculus
 FEATURE:
 OTHER INFORMATION: For all n's in this sequence, n=(a or g or c or t)
 US-08-841-349-15

Query Match 6.5%; Score 58.4; DB 2; Length 981;
 Best Local Similarity 49.8%; Pred. No. 8.3e-08;
 Matches 146; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 594 AGGATATGAGAGCCCAACAGGGGATATGAGAGCCCACTATGAGATATGAGAGCCCGCC 653
 DB 944 AGCCATATGAGCAGCTCTCCAGGCAATGATAGCAGCTTCCAGGCAATGATAGCAGCTC 885

QY 654 TGTGGATATGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATGAGTCCC 713
DB 884 TCCAGGCCATGTATGACAGCTCTCCAGGCCATGTATGACTCTCCAGGCTCATATGACGAG 825
QY 714 ACCCTGGGGATATGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATGAGTCC 773
DB 824 CTCTCAGAGCCATGTATGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATGAGTCC 765
QY 774 CCCACCTGGGGATATGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATGAGTCC 833
DB 764 CAGATCTCAGAGCCATGTATGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATGAGTCC 705
QY 834 AGCCCTACCCCTGATATGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATGAGTCC 886
DB 704 GGGCAGCTCTCAGAGCCATGTATGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATGAGTCC 652

RESULT 9
US-08-917-320-18
Sequence 18, Application US/08917320
Patent No. 5824508
GENERAL INFORMATION:
APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
TITLE OF INVENTION: No. 5824508 Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSES: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/917,320
FILING DATE: 25-AUG-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Luann Caert
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5163
TELEFAX: 415-857-0663
TELEX: 380816 CooleyPA
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1014..3734
US-08-917-320-18

Query Match 6.1%; Score 54.8; DB 1; Length 3833;
Best Local Similarity 48.7%; Pred. No. 2e-06;
Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
QY 516 AGCATATCCATTTGATCTATGAGCCCAACACAGAGATATACAGTCCACAGAGGGA 575
DB 2532 ACCAGCTCAGCTCAGAGCCATGTATGAGCCCAACCAAGCCAGCCAGCCAGCCAGGGA 2591

QY 576 ATATGAACTCCACAGAGATATGAGCCCAACAGGGGATATGAGCCCACTAT 635
DB 2532 GTGACTACCCCAACCCCAAAAGCCACAGCCCAACAGAGTACTACCCCAACCCA 2651
QY 636 GGGATATGAGCCCGCTGTGGATATGAGTCCACCTGGGGATATGAGTCCAC 695
DB 2652 AATGCCACAGCCCACTGTGGGAAACAAAGTCTTACAGAGTACCTACCCCAAC 2711
QY 696 TGGGGATATGAGTCCCACTGTGGGATATGAGCCCACTGTGGGGATATGAGTCC 755
DB 2712 CCAATATGCCACAGCCCACTGTGGGAAACAAAGTCTTACAGAGTACCTACCCCA 2771
QY 756 ACCCTGGGGATATGAGTCCCACTGTGGGATATGAGCCCACTGTGAGATATGAGAC 815
DB 2772 ACCCAATATGCCACAGCCCACTGTGGGAAACAAAGTCTTACAGAGTACCTACCCCA 2831
QY 816 CCCACC 821
DB 2832 CCAACC 2837

RESULT 10
PCT-US95-04611A-18
Sequence 18, Application PC/TUS9504611A
GENERAL INFORMATION:
APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
TITLE OF INVENTION: Non Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSES: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04611A
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Luann Caert
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5163
TELEFAX: 415-857-0663
TELEX: 380816 CooleyPA
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1014..3734
PCT-US95-04611A-18

Query Match 6.1%; Score 54.8; DB 5; Length 3833;
Best Local Similarity 48.7%; Pred. No. 2e-06;
Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
QY 516 AGCATATCCATTTGATCTATGAGCCCAACACAGAGATATACAGTCCACAGAGGGA 575

Db 2532 ACCAGCTCCACCTGACAGTACTACCCCAATGCCAAGCCCAAGCCCAAGCA 2591
Qy 576 ATATGAACTCCACCAAGAAATATGAGCCCAACAGGGGGATATGAGCCCACTAT 635
Db 2592 GTGACTACCCCAACCCCAATGACACCAAGCCCAAGCAAGTACTACCCCAAGCA 2651
Qy 636 GGAATATGAGCCCGCTGTGGGATATGAGTCCCACTGGGGATATGAGTCCACC 695
Db 2652 AATGCAACCAAGCCCACTGTGGGAAAAAACAAGCTCACTGAGAGTACTACCCCAAGC 2711
Qy 696 TGGGGGATATGAGTCCCACTGGGGGATATGAGCCCAAGCCCAAGTATGAGTCCC 755
Db 2712 CCAATGCAACCAAGCCCACTGTGGGAAAAAACAAGCCCACTGAGAGTACTACCCCA 2771
Qy 756 ACCTGGGGGATATGAGTCCCACTGGGGGATATGAGCCCAAGTATGAGTACC 815
Db 2772 ACCCAAAATGCAACCAAGCCCACTGTGGGAAAAAACAAGCCCACTGAGAGTACTACC 2831
Qy 816 CCCACC 821 0
Db 2832 CCAACC 2837

RESULT 11

US-08-783-774-1

Sequence 1, Application US/08783774

Patent No. 6054130

GENERAL INFORMATION:

APPLICANT: Spaete, Richard

APPLICANT: Jackman, Winthrop

TITLE OF INVENTION: NON-SPLICING VARIANTS OF

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESS: Penile & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036/2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/783,774

FILING DATE: 15-JAN-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7682-037

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-790-9090

TELEFAX: 212-869-8864

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5931 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 1014...3734

OTHER INFORMATION:

US-08-783-774-1

Query Match 6.1%; Score 54.8; DB 3; Length 5931;

Best Local Similarity 48.7%; Pred. No. 2.4e-06;

Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
Qy 516 AGCATATCCAAATTGATCTATGAGACCCCAACAGAGATATACGTCAACAGGGGA 575
Db 2532 ACCAGCTCCACCTGACAGTACTACCCCAATGCCAAGCCCAAGCCCAAGCA 2591
Qy 576 ATATGAACTCCACCAAGAAATATGAGCCCAACAGGGGGATATGAGCCCACTAT 635
Db 2592 GTGACTACCCCAACCCCAATGACACCAAGCCCAAGCAAGTACTACCCCAAGCA 2651
Qy 636 GGAATATGAGCCCGCTGTGGGATATGAGTCCCACTGGGGATATGAGTCCACC 695
Db 2652 AATGCAACCAAGCCCACTGTGGGAAAAAACAAGCTCACTGAGAGTACTACCCCAAGC 2711
Qy 696 TGGGGGATATGAGTCCCACTGGGGGATATGAGCCCAAGCCCAAGTATGAGTCCC 755
Db 2712 CCAATGCAACCAAGCCCACTGTGGGAAAAAACAAGCCCACTGAGAGTACTACCCCA 2771
Qy 756 ACCTGGGGGATATGAGTCCCACTGGGGGATATGAGCCCAAGTATGAGTACC 815
Db 2772 ACCCAAAATGCAACCAAGCCCACTGTGGGAAAAAACAAGCCCACTGAGAGTACTACC 2831
Qy 816 CCCACC 821
Db 2832 CCAACC 2837

RESULT 12

US-09-556-706B-1

Sequence 1, Application US/09556706B

Patent No. 6458364

GENERAL INFORMATION:

APPLICANT: Spaete, Richard

APPLICANT: Jackman, Winthrop

TITLE OF INVENTION: NON SPLICING VARIANTS OF GP350/220

FILE REFERENCE: 7682-050-999

CURRENT APPLICATION NUMBER: US/09/556,706B

CURRENT FILING DATE: 2000-04-24

PRIOR FILING DATE: 1997-01-15

PRIOR APPLICATION NUMBER: 08/783,774

PRIOR FILING DATE: 1994-04-18

NUMBER OF SEQ ID NOS: 19

SOFTWARE: PatentIn version 3.0

SEQ ID NO 1

LENGTH: 5931

TYPE: DNA

ORGANISM: Virus

FEATURE:

OTHER INFORMATION: GP350/220

US-09-556-706B-1

Query Match 6.1%; Score 54.8; DB 4; Length 5931;

Best Local Similarity 48.7%; Pred. No. 2.4e-06;

Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

Qy 516 AGCATATCCAAATTGATCTATGAGACCCCAACAGAGATATACGTCAACAGGGGA 575
Db 2532 ACCAGCTCCACCTGACAGTACTACCCCAATGCCAAGCCCAAGCCCAAGCA 2591
Qy 576 ATATGAACTCCACCAAGAAATATGAGCCCAACAGGGGGATATGAGCCCACTAT 635
Db 2592 GTGACTACCCCAACCCCAATGACACCAAGCCCAAGCAAGTACTACCCCAAGCA 2651
Qy 636 GGAATATGAGCCCGCTGTGGGATATGAGTCCCACTGGGGATATGAGTCCACC 695
Db 2652 AATGCAACCAAGCCCACTGTGGGAAAAAACAAGCTCACTGAGAGTACTACCCCAAGC 2711
Qy 696 TGGGGGATATGAGTCCCACTGGGGGATATGAGCCCAAGCCCAAGTATGAGTCCC 755
Db 2712 CCAATGCAACCAAGCCCACTGTGGGAAAAAACAAGCCCACTGAGAGTACTACCCCA 2771
Qy 756 ACCTGGGGGATATGAGTCCCACTGGGGGATATGAGCCCAAGTATGAGTACC 815

Db 2772 ACCCAATGACACACCCACCTTGGAAGAAACCCCACTGAGAGTACTACC 2831
Qy 816 CCCACC 821
Db 2832 CCAACC 2837

RESULT 13
US-08-182-175A-56
Sequence 56, Application US/08182175A
Patent No. 5559223
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,175A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axamethy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4829
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULAR TYPE: DNA (genomic)
ORIGINAL SOURCE:
CELL TYPE: B. coli
IMMEDIATE SOURCE:
CLONE: 2-9
FEATURE:
NAME/KEY: CDS
LOCATION: 2..235
OTHER INFORMATION: /function= "synthetic storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "sep"
OTHER INFORMATION: /standard_name= "7.7.7.7.7.8.9.8.9.5"
US-08-182-175A-56

Query Match 5.9%; Score 53.4; DB 1; Length 243;
Best Local Similarity 53.0%; Pred. No. 1.4e-06;
Matches 114; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 599 ATGGAGCCCAACGAGGGGATATGAGCCCACTATGAGTATGAGCCCGCTGTG 658
Db 2 ATGGAGGAGAGGCTGAAGCGCATGAGAGAGAGCTGAAGCGCATGAGAGAGCTGAAG 61

Qy 659 GATATGAGTCCACCTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 718
Db 62 GCGATGAGAGAGGCTGAAGCGCATGAGAGAGAGCTGAAGCGCATGAGAGAGCTG 121
Qy 719 GGGATATGAGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 778
Db 122 AAGCGATGAGAGAGAGGCTGAAGCGCATGAGAGAGAGCTGAAGCGCATGAGAGAGAG 181
Qy 779 CTGAGGATATGAGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 813
Db 182 CTCAGAGATGAGAGAGAGGCTGAAGCGCATGAGAGAGAGCTGAAGCGCATGAGAGAGAG 216

RESULT 14
US-08-474-633A-74
Sequence 74, Application US/08474633A
Patent No. 5773691
GENERAL INFORMATION:
APPLICANT: E. I. DU PONT DE NEMOURS AND
APPLICANT: COMPANY
TITLE OF INVENTION: CHIMERIC GENES AND
TITLE OF INVENTION: METHODS FOR INCREASING
TITLE OF INVENTION: INCREASING THE LYSINE
TITLE OF INVENTION: AND THRONINE CONTENT
TITLE OF INVENTION: OF THE SEEDS OF PLANTS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS
ADDRESSEE: AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,633A
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: BARBARA C. SIEGEL
REGISTRATION NUMBER: 30,684
REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4931
TELEFAX: 302-773-0164
TELEX: 835420
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULAR TYPE: DNA (genomic)
ORIGINAL SOURCE:
CELL TYPE: B. coli
IMMEDIATE SOURCE:
CLONE: 2-9
FEATURE:
NAME/KEY: CDS
LOCATION: 2..235
OTHER INFORMATION: /function= "synthetic
OTHER INFORMATION: storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "sep"
OTHER INFORMATION: /standard_name= "7.7.7.7.7.8.9.8.9.5"

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 10:43:10 / Search time 295.533 Seconds

(Without alignments)
10099.015 Million cell updates/sec

Title: US-09-864-291-4_COPY_36_933

Perfect score: 898
Sequence: 1 atggcagctgaccagcagcga.....tcgctctcacagatctatga 898

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2201672 seqs, 166179599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	262.6	29.2	467	11 US-09-764-891-1230	Sequence 1230, Ap
2	220	24.5	7099	11 US-09-764-891-1230	Sequence 1230, Ap
3	216.8	24.1	436	9 US-09-864-761-1561	Sequence 11561, A
4	195.4	21.8	321	11 US-09-764-891-1356	Sequence 1356, Ap
5	122	13.6	432	11 US-09-918-995-13536	Sequence 13536, A
6	114.4	12.7	409	10 US-09-960-352-6984	Sequence 6984, Ap
7	92.4	10.3	291	11 US-09-764-891-1571	Sequence 1571, Ap
8	86	9.6	471	9 US-09-864-761-15029	Sequence 15029, A
9	78.6	8.8	511	13 US-10-029-386-24981	Sequence 24981, A
10	77.8	8.7	593	13 US-10-027-632-290832	Sequence 290832, A
11	77.6	8.2	420	10 US-09-974-300-3626	Sequence 3626, Ap
12	73.6	8.2	866	13 US-10-029-386-22839	Sequence 22839, A
13	69.4	7.7	591	9 US-09-864-761-19727	Sequence 19727, A
14	68.8	7.7	1959	9 US-09-864-761-2946	Sequence 2946, Ap
15	68.8	7.7	1959	9 US-09-864-761-2946	Sequence 2946, Ap

16	68.6	7.6	2016	10 US-09-938-842A-2004	Sequence 2004, Ap
17	63.6	7.1	1040	13 US-10-029-386-26033	Sequence 26033, A
18	61.6	6.9	541	13 US-10-029-386-11265	Sequence 11265, A
19	59.8	6.7	3469	13 US-10-294-804-1	Sequence 1, Appl1
20	59.4	6.6	541	13 US-10-029-386-12333	Sequence 12333, A
21	58.8	6.5	417	9 US-09-864-761-27093	Sequence 27093, A
22	58.8	6.5	3048	13 US-10-027-632-113439	Sequence 113439, A
23	58.8	6.5	3048	14 US-10-027-632-113439	Sequence 113439, A
24	57.2	6.4	511	13 US-10-029-386-24981	Sequence 24981, A
25	55.6	6.2	628	13 US-10-029-386-22859	Sequence 22859, A
26	54.8	6.1	2108	10 US-09-962-832-225	Sequence 225, App
27	54.6	6.1	417	9 US-09-864-761-27093	Sequence 27093, A
28	54.6	6.1	422	10 US-09-738-973-337	Sequence 337, App
29	54.6	6.1	422	10 US-09-854-133-337	Sequence 337, App
30	54.6	6.1	422	15 US-10-144-642A-337	Sequence 337, App
31	54.6	6.1	3048	13 US-10-027-632-113439	Sequence 113439, A
32	54.6	6.1	3048	14 US-10-027-632-113439	Sequence 113439, A
33	53.8	6.0	541	13 US-10-029-386-11265	Sequence 11265, A
34	53.4	5.9	243	15 US-10-023-066A-74	Sequence 74, Appl
35	53.2	5.9	3773	9 US-09-925-302-47	Sequence 47, Appl
36	53.2	5.9	6354	15 US-10-084-817-158	Sequence 158, App
37	52.6	5.9	354	9 US-09-864-761-18194	Sequence 18194, A
38	52.6	5.9	431	13 US-10-029-386-14222	Sequence 14222, A
39	52.6	5.9	597	13 US-10-029-386-517	Sequence 517, App
40	52.6	5.9	2030	10 US-09-880-107-2416	Sequence 2416, Ap
41	52.6	5.9	8923	10 US-09-764-847-1041	Sequence 1041, Ap
42	52.6	5.9	8923	15 US-10-092-154-1041	Sequence 1041, Ap
43	52.4	5.8	2746	13 US-10-371-723-1	Sequence 1, Appl1
44	52.2	5.8	689	13 US-10-027-632-148124	Sequence 148124, A
45	52.2	5.8	689	14 US-10-027-632-148124	Sequence 148124, A

ALIGNMENTS

RESULT 1
US-09-764-891-1230
Sequence 1230, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PAM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1230
LENGTH: 467
TYPE: DNA
ORGANISM: Homo sapiens
FRAGMENT:
NAME/KEY: SITE
LOCATION: (421)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1230

Query Match 29.2% Score 262.6; DB 11, Length 467;
Best Local Similarity 76.4%; Pred. No. 2.1e-77;
Matches 365; Conservative 0; Mismatches 100; Indels 13; Gaps 3;
DB 351 CAGAGAGAGAGTGCATGATTCGCCAATCATGATGTAAGAGCTGCTCTGCTGC 410
2 CAGAAATGAGAGGTCATGATTCGCCAATCATGATGTAAGAGCTGCTCTGCTGC 61
CY 411 CAGAGATTCATCTGGAAGTGAATTAATTCGTTGACATTCAGAGCTGATCAAT 470
DB 62 CCAGAGATTCATCTGGAAGTGAATTAATTCGTTGACATTCAGAGCTGATCAAT 121
CY 471 TACTGTCACAGAGGCTGAGTGTCTTCAGAGACATTCGTCAGCATATTCAT 530
DB 122 TACTGTCACAGAGGAT---ATGTCACATTCAGAGATGCTTGTTCAG-----TTAT 169

QY 531 GATCTATGAGCCGCCACCAAGATATACAGTCCAAACAGGGGATATGAACTCCACC 590
DB 170 TGTCTATGAGGCCCACTGAGGATATGAGGCCCACTCCCGGATACAGAGCCCAACC 229
QY 591 AGAAGATATGAGGCCCACTGAGGATATGAGGCCCACTCTATGAGATATGAGGCCCC 650
DB 230 TGCAAGATATGAGGCCCACTGAGGATATGAGGCCCACTCTGAGATATGAGGCCCTC 289
QY 651 GCCTGTGAGATATGAGGCCCACTGAGGATATGAGGCCCACTCTGAGGATATGAGAGT 710
DB 290 ACTCTGTGAGATATGAGGCCCACTCTGAGATATGAGGCCCACTCTGAGATATGAGACC 349
QY 711 CCCACCTGAGGATATGAGGCCCACTCTGAGGATATGAGGCCCACTCTGAGGATATG 769
DB 350 CCCACCTCTGAGGATATGAGGCCCACTCTGAGATATGAGAACCCCACTCTGAGATATG 409
QY 770 GTGCCCCCACTGAGGATATGAGGCCCACTCTGAGGATATGAGGCCCACTCTGAG 827
DB 410 GAGCCCCCACTCTGAGATATGAGGCCCACTCTGAGGATATGAGGCCCACTCTGAGG 467

RESULT 2
US-09-764-891-7570
; Sequence 7570, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7570
; LENGTH: 7099
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7570

Query Match 24.5%; Score 220; DB 11; Length 7099;
Best Local Similarity 75.3%; Pred. No. 2,3e-62;
Matches 274; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 519 ATATCCATATGATCTATGAGCCGCCACCAAGATATACAGTCCAAACAGGGGATAT 578
DB 6718 ATTCCAGTATATGATCTATGAGGCCCACTCTGAGGATATGAGGCCCACTCTGAGAT 6777
QY 579 TGAAGCTCAGCAAGAGATATGAGGCCCAACAGGGGATATGAGGCCCACTATGAGG 638
DB 6778 CGGAGCCCACTGAGGATATGAGGCCCACTCTGAGGATATGAGGCCCACTCTGAGG 6837
QY 639 ATATGAGCCGCCCTGAGGATATGAGGCCCACTCTGAGGATATGAGGCCCACTCTG 698
DB 6838 ATACAGAGCTCAGCTGAGGATATGAGGCCCACTCTGAGATATGAGGCCCACTCTG 6897
QY 699 GGGATATGAGGCCCACTCTGAGGATATGAGGCCCACTCTGAGGATATGAGGCCCACT 758
DB 6898 AGGATATGAGGCCCACTCTGAGGATATGAGGCCCACTCTGAGATATGAGAACCCCACT 6957
QY 759 TGGGGGATATGAGGCCCACTCTGAGGATATGAGGCCCACTCTGAGGATATGAGGCCCC 818
DB 6998 TCTGAGATATGAGGCCCACTCTGAGGATATGAGGCCCACTCTGAGGATATGAGGCCCC 7017
QY 819 ACCAGCTGAGGATATGAGGCCCACTCTGAGGATATGAGGCCCACTCTGAGGATATGAG 878
DB 7018 GCTCTGAGGATATGAGGCCCACTCTGAGGATATGAGGCCCACTCTGAGGATATGAG 7077
QY 879 TGCC 882
DB 7078 AGCC 7081

RESULT 3
US-09-864-761-31561
; Sequence 31561, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31561
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Homo sapiens

FEATURE:
OTHER INFORMATION: MAP TO 299716.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
OTHER INFORMATION: BEST HUMAN HIT: AM169980.1, EVALU8 8.00e-11
OTHER INFORMATION: SWISSPROT HIT: P18616, EVALU8 3.00e-13
US-09-864-761-31561

Query Match 24.1%; Score 216.8; DB 9; Length 436;
Best Local Similarity 75.6%; Pred. No. 5.2e-62;
Matches 269; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 527 TTGATATATGAGCCGCCACCAAGATATACAGTCCAAACAGGGGATATGAACTC 586
|||||

Db 1 TTATGTCTATGAGCCCACTGACAGATATGAGCCCACTCCGATACGAGCCC 60
Qy 587 CACCAAGAGATATGAGCCCAACAGAGGATATGAGCCCACTATGAGATATGAG 646
Db 61 CACTGAGATATGAGCCCAACAGAGGATATGAGCCCACTATGAGATATGAG 120
Qy 647 CCCGACCTGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 706
Db 121 CTTACCTGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 180
Qy 707 GAGTCCCACTGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 766
Db 181 GAGCCCACTGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 240
Qy 767 ATGTGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 826
Db 241 ATGAGAGCCCACTGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 300
Qy 827 GAATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 882
Db 301 GATACAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 356

RESULT 4

US-09-764-891-1356
; Sequence 1356, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIORITY FILING DATE: 2001-01-17
; PRIORITY FILING DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1356
; LENGTH: 321
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (316)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (321)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1356

Query Match 21.8%; Score 195.4; DB 11; Length 321;

Best Local Similarity 77.4%; Pred. No. 6.8e-55;

Matches 246; Conservative 2; Mismatches 69; Indels 1; Gaps 1;
Qy 29 GCGGTGAGGAGCCCTGATCCCTGAGGAGAGATGCTTGAAGACATGAGAGATGAG 88
Db 1 GCGGCCGAGAGCCCTGATCCCTGAGGAGAGATGCTTGAAGACATGAGAGATGAG 60
Qy 89 ACCTGCTCTCTAAGAGAGAGAGATGCTTGAAGACATGAGAGATGAGAGAGAGAG 148
Db 61 AGCTCTCTCTCTAAGAGAGAGAGATGCTTGAAGACATGAGAGATGAGAGAGAGAG 120
Qy 149 CGTTGTTCTCTAAGAGAGAGAGATGCTTGAAGACATGAGAGATGAGAGAGAGAG 208
Db 121 CATGTGTTCTCTAAGAGAGAGAGATGCTTGAAGACATGAGAGATGAGAGAGAGAG 180
Qy 209 TGCTTCTCTAAGAGAGAGAGATGCTTGAAGACATGAGAGATGAGAGAGAGAGATG 268
Db 181 TGTTGTTCTCTAAGAGAGAGAGATGCTTGAAGACATGAGAGATGAGAGAGAGAGAT 240
Qy 269 TTGCCCCCACTAAGAGAGAGAGATGCTTGAAGAGAGAGAGAGAGAGAGAGAGAG 327
Db 241 TTGCTGAGAACTTATTAAGAGAGAGAGATGCTTGAAGAGAGAGAGAGAGAGAGAGAG 300

Qy 328 CAAGCTGTTTAAAGTAA 345
Db 301 CAAGCTACTTTAAATTA 318

RESULT 5

US-09-918-995-33536
; Sequence 33536, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIORITY FILING DATE: 2001-07-30
; PRIORITY FILING DATA: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: PatSeq for Windows Version 3.0
; SEQ ID NO 33536
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(432)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-33536

Query Match 13.6%; Score 122; DB 11; Length 432;

Best Local Similarity 59.4%; Pred. No. 4e-30;

Matches 247; Conservative 0; Mismatches 160; Indels 9; Gaps 2;
Qy 1 ATGCAAGTGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Db 18 ATGCGCGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 74
Qy 61 AGTGTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 75 AGCATCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 134
Qy 121 TATCTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Db 135 GAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 194
Qy 181 GTGACTTCACTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 195 CTGTC-----CAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 248
Qy 241 AGTACTGACATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 249 AAGAGCTGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 308
Qy 301 GCACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 309 GCGAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 368
Qy 361 GTGAGCATGAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 416
Db 369 GCGGCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424

RESULT 6

US-09-960-352-6984
; Sequence 6984, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathias, Naganan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 6984
LENGTH: 409
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 30-LIB3058-024-01-K1-H9
US-09-960-352-6984

Query Match 12.7%; Score 114.4; DB 10; Length 409;
Best Local Similarity 59.4%; Pred. No. 1.4e-27;
Matches 234; Conservative 0; Mismatches 151; Indels 9; Gaps 2;

QY 1 ATGGAGTGAACAGACGACACCGAGCGCTGCTGCGGCGCTCATCCCTCTGGCGAA 60
DB 23 ATGGCTCATCAAGAACCACTCGAGGCGCGGAGTGATGTCACAAACAC--CGAG 79
QY 61 AGTGTCTTGAAGCATGTGAGAGATGTGACCTCTGCTCTTACAGAAACAGTGAATCC 120
DB 80 AGCATCTTGAAGTCTTATGACCATGTAGAGCTTACGTTACATGACATGAGAAATGCCA 139
QY 121 TATCTTATATGACAAAGAAAGAAAGTGTCTTCTCATCTTACCTGCGTGTCTTC 180
DB 140 GAGGCTTCAAGGAGCAAGAAAGCAACCGTCTTACCTTACCCGCTACCGGTCTCTT 199
QY 181 GTGACTTCACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
DB 200 CTGTC-----CAAGCGAGGAGATGATGATGATGATGATGATGATGATGATGATG 253
QY 241 AGTGTCTTGAAGCATGTGAGAGATGTGACCTCTGCTCTTACAGAAACAGTGAATCC 300
DB 254 AAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 313
QY 301 GAGGCTTCAAGGAGCAAGAAAGCAACCGTCTTACCTTACCCGCTACCGGTCTCTT 360
DB 314 GCCGAAAGCAAGGAGTGTGCGGAGGAGGCTGTGATGATGATGATGATGATGATGATG 373
QY 361 GGTGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 394
DB 374 GCGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 407

RESULT 7
US-09-764-891-7571
Sequence 7571, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7571
LENGTH: 291
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-891-7571

Query Match 10.3%; Score 92.4; DB 11; Length 291;
Best Local Similarity 76.0%; Pred. No. 2.9e-20;
Matches 114; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 171 GGTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 230
DB 9 GGTGATTTTCATTAATTCATGCTCCATGATGATGATGATGATGATGATGATGATGATG 68
QY 231 TGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 290

DB 69 TGATCTATGACGAACTCTGATGAAACAGATATTTGTCGAACATTCATTAAGGG 128
QY 291 AACCATCAGGACGCTCCAGGTGTGCTG 320
DB 129 AACTATTCAGGACGCTCCATATGATGATG 158

RESULT 8
US-09-864-761-15029
Sequence 15029, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wenheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aesomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 15029
LENGTH: 471
TYPE: DNA
ORGANISM: Homo sapiens

FEATURE:
OTHER INFORMATION: MAP TO 299716.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
US-09-864-761-15029

Query Match 9.6%; Score 86; DB 9; Length 471;
Best Local Similarity 76.0%; Pred. No. 2.9e-20;
Matches 114; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Best Local Similarity 77.6%; Pred. No. 5,4e-18;
Matches 104; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 627 CCCACCTATGAGATATGAGAGCCCGCTGTGGATATGAGTCCACCTGGGGATATGAG 666
DB 338 CCCAGTATATGCTATATGAGAGCCCGCTGTGGATATGAGTCCACCTGGGGATATGAG 397
QY 687 AGTCCCACTGGGGATATGAGATGCTGAGGAGATATGAGAGCCCGCTGTGGATATGAG 746
DB 398 AGCCCACTGGAGATATGAGAGCCCGCTGTGGATATGAGAGCCCGCTGTGGATATGAG 457
QY 747 TGAAGTCCCACTG 760
DB 458 CAGAGCTCACTG 471

RESULT 9

US-10-029-386-24981
Sequence 24981, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AROMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 24981
LENGTH: 511
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO 297205.1
US-10-029-386-24981

Query Match 8.8%; Score 78.6; DB 13; Length 511;

Best Local Similarity 49.9%; Pred. No. 1.8e-15;
Matches 198; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

QY 485 CTGCACTGCTCTCTCAAGACACTTGTCCAGATATCCAAATGTGATCTATGAGCCC 544
DB 106 CTGCAAGGCCCCCTCTCTCAAGATGCTCCCTCTCTCAAGGAGGCTCCCTCTCTC 165
QY 545 CACCAACAGATATGAGTCCAAACAGAGGATATGAGACTCCACCAAGATATGAGAG 604
DB 166 CACTGAGGCTCCCTCTCTCAAGTCCAGGCTCCCTCTCTCAAGGCTCCCTCTC 225
QY 605 CCAACAGAGGATATGAGAGCCCACTATGAGATATGAGCCCGCTGTGGATATGAG 664
DB 226 CTCACTGAGGCTCTCTCTCTCTCAAGTCCAGGCTCCCTCTCTCAAGGCTCTCT 285
QY 665 GAGTCCCACTGGAGATATGAGATGCTCCAGCTGGAGATATGAGTCCCACTGGAGAT 724
DB 286 CTCTCACTGAGGCTCTCTCTCTCTCTCAAGGCTCTCTCTCTCAAGGCTCTCT 345
QY 725 ATGAGAGCCCACTGGAGATATGAGATGCTCCAGCTGGAGATATGAGTCCCACTGGAG 784
DB 346 CCTCTCTCTCACTGAGGCTCTCTCTCTCTCTCAAGGCTCTCTCTCTCAAGGCTCT 405
QY 785 GATATGAGAGCCCACTGGAGATATGAGAGCCCAAGCTGGAATATGAGAGCCCTACCCC 844
DB 406 GCTCCCTCTCTCACTGAGGCTCTCTCTCTCTCAAGGCTCTCTCTCTCAAGGCTCT 465
QY 845 CTGCAATGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 881
DB 466 CAGGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 502

RESULT 10

US-10-027-632-290832
Sequence 290832, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 290832
LENGTH: 593
TYPE: DNA
ORGANISM: Human
US-10-027-632-290832

Query Match 8.7%; Score 77.8; DB 13; Length 593;

Best Local Similarity 54.9%; Pred. No. 3.6e-15;
Matches 151; Conservative 1; Mismatches 123; Indels 0; Gaps 0;

QY 142 AAGGAACTGTCTCTCACTTATCCGGAGTCTTGTGACTTCACTTATGATCAT 201
DB 45 AAGGAACTGTCTCTCTCACTTATCCGGAGTCTTGTGACTTCACTTATGATCAT 104
QY 202 GACCCAGCTTCTTTTATGATGAGGCTTGTGAGTATGAGTATGAGTATGAGTATGAG 261
DB 105 GATGCAAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 164
QY 262 CCAATTTTGTCCCACTTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAG 321
DB 165 CCGTGTGAGCAAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAG 224
QY 322 GAAAGCAAGTCTTTTATGATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAG 381
DB 225 GAAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 284
QY 382 CTGATGTATGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 416
DB 285 CAGGTCTCTCAAGTATGATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAG 319

RESULT 11

US-10-027-632-290832
Sequence 290832, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483

;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 290832
;; LENGTH: 593
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-290832

Query Match 8.7%; Score 77.8; DB 14; Length 593;
Best Local Similarity 54.9%; Pred. No. 3.6e-15;
Matches 151; Conservative 1; Mismatches 123; Indels 0; Gaps 0;

QY 142 AAGAGAGCTTTCTGACCTTACCGGGGTCTTGTGACCTTCACTTAGTCAAT 201
DB 45 AAGGAGACCAAGAAAGGACCACTTACCTTCTTCTGTCAGGCGAAG 104
QY 202 GACCCAGCTTTCTTTATGATGCGGTTGGCTGATGATGATGACCACTTGAACA 261
DB 105 GATGCAAGCACTCTTCTGATATGATTTATCTTGAAGGCTATGATGATCAAGCA 164
QY 262 CCAATTTTTCCTCCCACTACTTAAGAAACCAATTCAGGCGCTCGAGTGTGGCTGG 321
DB 165 CCGTGTGTGACAACTGATCAAGGGAACAGTGAACCAAGGCGGTGTGCTGG 224
QY 322 GAAGGAGAGCTTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 381
DB 225 GAAGGCTCTGCTTCTGCAAGTCCGCTTCAAGGCGGCGGCGGCTGATGATGATG 284
QY 382 CTGATGTAAAGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 416
DB 285 CAGGTCTCTCAAGTGAATCTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGG 319

RESULT 12
US-09-974-300-3626
; Sequence 3626; Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/580,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3626
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-3626

Query Match 8.2%; Score 73.6; DB 10; Length 420;
Best Local Similarity 50.3%; Pred. No. 7.5e-14;
Matches 181; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
QY 522 TCGAATGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 581
DB 29 TCGAAGCGGCTGATATGATGATGATGATGATGATGATGATGATGATGATGATG 88

QY 582 AACTCCACAGAGATATGAGCCCAACGAGGGGATATGAGCCCCCACTATGAGATA 641
DB 89 CAGTCGAAGCGGCTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 148
QY 642 TGGAGCCCCGCTTGTGGGATATGAGTCCCACTGGGGGATATGAGATCCCACTGGGG 701
DB 149 TGGAGTCCGAAGCGGCTGATATGATGATGATGATGATGATGATGATGATGATGAT 208
QY 702 ATATGAGTCCCACTGGGGGATATGAGCCCACTGGGGGATATGAGATCCCACTGG 761
DB 209 ATATGAGTCCCACTGGGGGATATGAGCCCACTGGGGGATATGAGATCCCACTGG 268
QY 762 GAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 821
DB 269 CTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 328
QY 822 AGCTGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 881
DB 329 CCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 388

RESULT 13
US-10-029-386-22839/C
; Sequence 22839; Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DETERMINED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22839
; LENGTH: 866
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; OTHER INFORMATION: MAP TO AC005630.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: SWISSPROT HIT: 008379, EVALU 4.00e-10
; OTHER INFORMATION: NT HIT: g16161055, EVALU 4.00e-95
; OTHER INFORMATION: EST_HUMAN HIT: BF828725.1, EVALU 0.00e+00
US-10-029-386-22839

Query Match 7.7%; Score 69.4; DB 13; Length 866;
Best Local Similarity 49.9%; Pred. No. 3e-12;
Matches 175; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 536 ATGAGCCCCCAACAGATATATGATGATGATGATGATGATGATGATGATGATGAT 595
DB 509 AGTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 450
QY 596 GATATGAGCCCAACAGGGGATATGAGCCCACTGATGATGATGATGATGATGATGAT 655
DB 449 TTCAACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 390
QY 656 TGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 715
DB 389 CTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 330
QY 716 CTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 775
DB 329 CTCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 270
QY 776 CACTGGGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 835

Db 269 CTCCTCCGTTCACGTAGCTCTCTCTGTTACAGAGCTCTCTCTGTCATGATAG 210
QY 836 CCCATCCCCGTCATATGAAGTCATCTGTGAAATAGCTGCTCTC 886
Db 209 CTCCTCCGTTCATGATAGCTCTCTCTGTTACAGAGCTCTCTC 159

RESULT 14

US-09-864-761-19727
Sequence 19727, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 19727
LENGTH: 591
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
OTHER INFORMATION: MAP TO AC002041.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.8
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.2

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.3
OTHER INFORMATION: SWISSPROT HIT: P03200, EVALU8 3.00e-16
US-09-864-761-19727

Query Match 7.7%; Score 68.8; DB 9; Length 591;
Best Local Similarity 51.6%; Pred. No. 3.8e-12;
Matches 157; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 567 ACCAGGGGAATATGAACTCCACCAAGATATGAGCCCAACGAGGGGATATGAGC 626
Db 170 ACCGAGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 229
QY 627 CCCACCTATGAGATATGAGAGCCCCCTGATGATATGAGTCCACCTGAGGGATATG 686
Db 230 GTACCTCTGTGACATGTGATGATGATGATGATGATGATGATGATGATGATGATG 289
QY 687 AGTCCCACTGAGGGGATATGAGATCCCACTGAGGGGATATGAGCCCACTGAGGGGATA 746
Db 290 GAGGTACCTGAGGGGATATGATGATGATGATGATGATGATGATGATGATGATGATG 349
QY 747 TGAATGCCACCTGAGGGGATATGATGATGATGATGATGATGATGATGATGATGATG 806
Db 350 TGTAGAGTACGATGAGGGGATATGATGATGATGATGATGATGATGATGATGATGATG 409
QY 807 ATATGAGAGCCCACTGAGGGGATATGAGATCCCACTGAGGGGATATGAGATCCCACTG 866
Db 410 ATGTGAGAGTACCTGAGGGGATATGATGATGATGATGATGATGATGATGATGATGATG 469
QY 867 TGA 870
Db 470 GGGA 473

RESULT 15

US-09-864-761-2946
Sequence 2946, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662

;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 2946
;; LENGTH: 1959
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC002041.1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.4
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.8
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.9
;; OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 4.2
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.2
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.3
US-09-864-761-2946

Query Match 7.7%; Score 68.8; DB 9; Length 1959;
Best Local Similarity 51.6%; Pred. No. 7.7e-12;
Matches 157; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
QY 567 ACCAGGGGATATGGAATCTCCACGAGAGATATGAGCCCAACGAGGGGATATGAGC 626
DB 479 ACCTGGGGTATGTGAGGTGACGTGGGGAATGTGTGAGGTGACCTGGGGGACGTGTGAG 538
QY 627 CCCACTATGGGATATGAGAGCCCGCTGGGATATGAGTCCCACTGGGGGATATGG 686
DB 539 GTGACCTGTGGACATGTGAGGTGACCTGGGGGATGTGTGAGGTGACCTGGGGGACGTGT 598
QY 687 AGTCCACCTGGGGATATGAGTCCCACTGGGGATATGAGCCCACTGGGGGATATG 746
DB 539 GAGGTGACCTGGGGGATGTGTGAGGTGACCTGTGGGACATGTGAGGTGACGTGGCAATG 658
QY 747 TGAAGTCCCACTGGGGGATATGTGCCCACTGGGGGATATGAGCCCACTGCAAG 806
DB 659 TGTGAGGTGACGTGGGGAATGTGTGAGGTGACCTGGGGGACGTGTGACGTGGGGA 718
QY 807 ATATGAGAGCCCAACGAGCTGGAATAATGAAGCCCTACCCCTGCATATGAAGCTCCATCTGC 866
DB 719 ATGTGTGAGGTGACCTGGGGGAACGTGTGAGGTGACCTGGGGGATGTGTGAGGTGACCTGG 778
QY 867 TGA 870
DB 779 GGA 782

Search completed: December 16, 2003, 17:58:43
Job time : 296.533 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: December 16, 2003, 09:21:45 ; Search time 1848.09 seconds
(without alignments)
11809.735 Million cell updates/sec

Title: US-09-864-291-4_COPY_36_933

Sequence: 1 atgcgcagtgaccagagccca.....tgcctctcacagatcctatga 898

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estrA:*
2: em_estrhum:*
3: em_estrin:*
4: em_estrnu:*
5: em_estrsv:*
6: em_estrpl:*
7: em_estrro:*
8: em_estrh:*
9: gb_estr1:*
10: gb_estr2:*
11: gb_estr3:*
12: gb_estr4:*
13: gb_estr5:*
14: gb_estr6:*
15: em_estrfun:*
16: em_estrcom:*
17: em_gse_hum:*
18: em_gse_inv:*
19: em_gse_pln:*
20: em_gse_vrt:*
21: em_gse_fun:*
22: em_gse_mam:*
23: em_gse_mus:*
24: em_gse_pro:*
25: em_gse_rtd:*
26: em_gse_pbg:*
27: em_gse_vrt1:*
28: gb_gse1:*
29: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	505.8	56.3	1079	12	BM564167 AGENCOURT
2	414.4	46.1	1414	11	AK015863 Mus muscu
3	340.8	38.0	663	10	BG699398 602679081
4	340	37.9	775	10	BG722815 602695278

5	333.8	37.2	717	14	BY715414
6	206	22.9	701	10	BG701881
7	189.8	21.1	360	13	BY098622
8	183.6	20.4	932	13	BU107273
9	179.4	20.0	560	13	BU103740
10	179.4	20.0	700	13	BU284977
11	179.4	20.0	723	9	AJ452591
12	177.8	19.8	568	12	BT390687
13	177.8	19.8	636	12	BM490729
14	167.4	18.6	772	9	AJ454950
15	166.8	18.6	706	13	BU433499
16	153.6	17.1	816	13	BU168990
17	153.6	17.1	931	13	BU914288
18	152.2	16.9	613	9	AL849211
19	152.2	16.9	647	9	AL849651
20	152.2	16.9	655	9	AL988723
21	152.2	16.9	661	9	AL972453
22	152.2	16.9	663	9	AL881810
23	152.2	16.9	663	9	AL896682
24	152	16.9	24	10	BR680329
25	152	16.9	509	10	BS15416
26	152	16.9	594	12	BM190972
27	152	16.9	607	12	BM190975
28	152	16.9	868	14	CD255258
29	151.6	16.9	460	10	BS513187
30	151	16.8	719	14	CS593075
31	150.6	16.8	611	9	AL846063
32	150.6	16.8	760	14	CS592279
33	150.4	16.7	726	14	CS591559
34	150.4	16.7	862	13	BO735296
35	150.4	16.7	892	14	CD301781
36	143.6	16.0	488	28	AO997933
37	141.2	15.7	635	9	AL889954
38	137.2	15.3	814	13	BU961805
39	136.8	15.2	622	10	BR468886
40	136.6	15.1	899	10	BR470466
41	132.8	14.8	652	14	CA590725
42	132.8	14.7	631	13	BU805447
43	131.8	14.7	744	13	BU230211
44	131.2	14.6	635	13	BU807307
45	130	14.5	437	13	BY249862

ALIGNMENTS

RESULT 1
LOCUS BM564167 1079 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6560149 NIH_MGC_119 Homo sapiens CDNA clone IMAGE:5742129
ACCESSION BM564167
VERSION BM564167.1 GI:18811738
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1079)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12759 row: m column: 10
High quality sequence stop: 726.

FEATURES
Source

Location/Qualifiers
1. 1079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5742129"
/tissue_type="medulla"
/lab_host="DH10B"
/clone_id="NIH_MGC_119"
/note="Organ: brain; Vector: pCMV-Sport6; Site: 1; Not; Site 2: EcorV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dt primed and directionally cloned (EcorV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."

BASE COUNT 267 a 297 c 268 g 247 t

ORIGIN

Query Match 56.3%; Score 505.8; DB 12; Length 1079;
Best Local Similarity 75.7%; Pred. No. 1.4e-135;
Matches 670; Conservative 0; Mismatches 202; Indels 13; Gaps 3;

1 ATGGCAGTGAACCAAGCCACACCGAGCGCTGCGGGCCCTCATCCCTCTGCGCAA 60
2 |||||
3 47 AAGGCGGTGAATCAGAGCCACACGAGAACCGCGCGGAGCCCTCATCCCTTAAGGTGAA 106
4 |||||
5 61 AGTGTCTTGAACGAGTGTGAGATGTGACCTCTGCTCTCAACAGAAACGATGGAATCC 120
6 |||||
7 107 AGTCTCTTGAAGCGGTCTCCAAATGTGAGCTCTCTCCACAGCATCAGAAAGCTCA 166
8 |||||
9 121 TATCTCTTAAAGCAGAAAGAAAGAGCTTTTCTCATCTCAATACCGGTGCTTC 180
10 |||||
11 167 AATGCTTTAGTGTGAAGACAGAAACATTTGTTCTCATCTCAATACCGGTGATTTTC 226
12 |||||
13 181 GTGACTTCACTTATGTCATATGACCCCAATGCTTTCTTTATGATGCCCTTTG 240
14 |||||
15 227 AATACCTTATGCTCATGATGATCCCATGTTGCTTTTATGATGCCATTTGATCTATG 286
16 |||||
17 241 AGTACGTGACCATTTGAACAAACCAATTTTGGCCCACTACATTAAGAAACCATTCAG 300
18 |||||
19 287 ACGAACCTCAGCTGTGAACAAACAGTATTTCTGCAACCTCATTAAGGAACTATTCAG 346
20 |||||
21 301 GCACTTCAAGTGTGCGTGGAGAGAAAGCTGTTTAACTTATCTTCAAGAAAGCA 360
22 |||||
23 347 GCACTTCAATGTGTGCGTGGAGAGAAAGCTACTTTTAAATTAAGTTCAAGAAATGCA 406
24 |||||
25 361 GATGCGATCATTTTGGCCCACTGATGTGTAAGGTGCTCTGCTGCTGCGAGAGAAAT 420
26 |||||
27 407 GTGTGCTTGAATTTGGCCCACTGATGTGTAAGGTGCTCTGCTGCTGCGAGAGATTT 466
28 |||||
29 421 CCACTTGAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 480
30 |||||
31 467 CCACTTGAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 526
32 |||||
33 481 GGGGCTGAGTGTGCTCTCAACGACCTTGTCCAGATATCCAAATGTGATCTATGGA 540
34 |||||
35 527 GGGAAAT--ATGTGACTTCAACAGATGCTTGTGAG-----TATATGTCTATGGA 574
36 |||||
37 541 CCCCCACCAAGATATATACAGTCCAAACAGGGGAAATATGAACTCCACAGAAAGATAT 600
38 |||||
39 575 GCCCACTGAGAGATATGAGGCCCACTCCCGGATACGAGAGCCCACTGAGAGATAT 634
40 |||||
41 601 GAGGCCCAACAGGGGGATATGAGGCCCACTATGAGATATGAGGCCCGCTGTGGA 660
42 |||||
43 635 GAGGCCCAACCGTGTGAGAAATGAGGCCCGCTGTGAGATATGAGGCCCTCATCTGTGGA 694
44 |||||
45 661 TATGAGTCCCACTGGGGATATGAGATCCCACTGGGGATATGAG--TCCACCTGG 719
46 |||||
47 695 TATGAGGCCCACTGTGAGATACGAGGCCCACTGAGATATGAGGCCCACTCT 754

QY 720 GGGATATGAGGCCCACTGGGGATATGAGATCCCACTGGGGATATGATGCCCAACC 779
DB 755 AGATATGAGAGCCCACTGCTGATATGAGAACCCCACTCTGATATGAGAGCCCAACC 814
QY 780 TGGGGATATGAGAGCCCACTGCTGATATGAGAGCCCACTGCTGATATGAGAGCCCT 839
DB 815 TCTGGATATGAGAGCCCACTGCTGATATGAGAGCCCACTGCTGATATGAGAGCCCTC 874
QY 840 ACCCGCTGATATGAGAGCCCACTGCTGATATGAGAGCCCACTGCTGATATGAGAGCCCTC 884
DB 875 ACCGCTGATATGAGAGCCCACTGCTGATATGAGAGCCCACTGCTGATATGAGAGCCCTC 919

RESULT 2
AK015863
LOCUS
DEFINITION
AK015863 1414 bp mRNA linear HTC 05-DEC-2002
Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:149305211r23 product:hypothetical protein, full insert sequence.
AK015863
VERSION
AK015863.1 GI:12854363
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636

REFERENCE
AUTHORS
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE
Normalization and subcloning of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE
20499374
PUBMED
11042159

REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kizumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaue, S., Ikegami, T., Kashiwagi, K., Fujisawa, S., Inoue, K., Togawa, Y., Iizawa, K., Tanaka, T., Matsura, S., Kawai, J., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.
TITLE
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE
20530913
PUBMED
11076861

REFERENCE
AUTHORS
Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, K., Iizawa, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakawa, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, M., Gaasterland, T., Glisic, C., King, B., Kochiwa, H., Knehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Penzo, G., Quackenbush, J., Schiraldi, L. M., Seabolt, P., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barab, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bona, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. P., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilmink, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohseki, S., and Hayashizaki, Y.

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
BG699398	602679081P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4811804 5', mRNA sequence.	BG699398	1	GI:13967653	EST.	Homo sapiens (human)	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 653)	NIH-MGC http://mgc.nci.nih.gov/ .	National Institutes of Health, Mammalian Gene Collection (MGC)
							Unpublished	Contact: Robert Strausberg, Ph.D.			Email: cgabbs-remail.nih.gov
								Tissue Procurement: Miklos Pavlovits, M.D., Ph.D.			
								cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki Toshitaki and Piero Carninci (RIKEN)			
								cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNLN)			
								DNA Sequencing by: Incyte Genomics, Inc.			
								Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:			
								http://image.llnl.gov			
								Plate: LLM10703 row: a column: 21			
								High quality sequence stop: 660.			
								Location/Qualifiers			
								1..663			

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4811804"
/tissue_type="hippocampus"
/lab_note="DH10B"
/clone_idb="NIH_MGC_95"
/note="Organ: brain; Vector: pluescript (modified
pluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gcgcag
); Oligo-dT primed using primer 5'-TCTTTTCTTTTCTTTT-3',
size-selected for average insert size 2.5 kb and
normalized to RQF 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

```

```

Query Match Similarity 38.0%; Score 340.8; DB 10; Length 663;
Best Local Similarity 75.0%; Pred. No. 9, 8e-88;
Matches 481; Conservative 0; Mismatches 145; Indels 15; Gaps 4

QY      1  ATGCAAGTGAACCAAGCCACACCCAGAGCCGCTCGTGGGACCTCATCCCTCTGGCCGA 60
      |||||
DB      37  ATGGCGGTGAATCAAGGCCACACCCGGAACCCCGGGAGCCCTCATCCCTAACGGTGA 96
      |||||

QY      61  AGTGTCTGAACCAAGTGGAGATGTGGACCTGTGCTCTTACAGAAACAGTGGAAATCC 120
      |||||
DB      97  AGTCTCTTGAAACGGGCTCCGAAATGTGAGCTCTCTTCCACAGCATCAGAAAGCTCA 156
      |||||

QY      121  TATCTCTTAAATGGCAACAAGAAAGGAAGCTTGTTCCTCACTTCATACCGGGTGTCTTC 180
      |||||
DB      157  AATGCTTTTAGTGTAGAAAGACAAGGAC-ATGTTCTCACTTCATACCGGGTGAATTTTC 215
      |||||

QY      181  GCGACTTCGCACTTATGTCAATGACCCCAAGCTTCTTTATATGATCCGTTTGGCCGATG 240
      |||||
DB      216  ATAACCTCATGCTCCATCAAGTATCCCATGATGCTTTTATATATGCAATTTATCTGATG 275
      |||||

QY      241  AGTGACTGCACCATTTGAACAACAATTTTGTGCCCCAATCAATTAAGGAACCAATTGAG 300
      |||||
DB      276  ACGAAGCTCACTGTTGAACAACAAGTAATTTGTGCGAAATCTTAATTAAGGAACATAATGAG 335
      |||||

QY      301  GCAAGTCCAGGTGTGGCTGGGAAGCAAGCTGTTTAAATTAAGTTATCTTTCAGAAAGGA 360
      |||||

```

Db	336	GCAGCTCCATATGGTGGCTGGGGAAGGACACCTACTTTTAAATTAAGCTTCAGAAATGGA	335
Qy	361	GGAGCATGGAAATTTGGCCCACTGATGATTAAGCTGCTCTGCTGCGCAGAGGAATT	420
Db	396	GGATCATTTGAATTTTGCCCATGTTATGATGAAAGCTGCTCTGCTGCGCCAGAGATTT	455
Qy	421	CCACTTGAAGTGTAAATTACTGTTTGACACTTCAGAGACTGTACATTAATTACTGTGCCA	480
Db	456	CCACTTGAAGCTTAAATGAACCTGGTTCAAGCTCTATGGAATTTATGTAAATTAACCTGGGGA	515
Qy	481	GGGGCTGCAGTGTGCTCTCAACAGACACTTGTCCAGCATATTCAAATTGTGATCTTATGGA	540
Db	516	GGGAAT---ATGTGCATCTCCACAGATGCTTGTTCAG-----TTATGTGTATGGA	553
Qy	541	CCCCCAACACAGATATTAACGTCCAAACAGGGGAATATGGAACCTCCACAGAAAGATAT	600
Db	564	GGCCCACTGCAAGGATATGAGCCCCCACTCCCGGA--TACGAGACCCCACTGCAAGATAT	621
Qy	601	GGAGCCCAACAGGGGGGATATGAGACCCCACTTATGGGATAT	641
Db	622	GGAGCCCAACCTATAGAAATATAGACGCCGCTGTGGGATAT	662

RESULT	4
BG722815	
LOCUS	BG722815
DEFINITION	BG722815 775 bp mRNA linear EST 08-MAY-2001
ACCESSION	60269592781 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4827427 5' ,
VERSION	BG722815
KEYWORDS	BG722815.1 GI:14002002
SOURCE	EST.
ORGANISM	Homo sapiens (human)
	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
				Unpublished
				Contact: Robert Strausberg, Ph.D.
				Email: cga@dc-remail.nih.gov
				Tissue Procurement: Miklos Pavlovits, M.D., Ph.D.
				cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
				Toshiyuki and Piero Carninci (RIKEN)
				cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLT)
				DNA Sequencing by: Incyte Genomics, Inc.
				Clone distribution: WGC clone distribution information can be
				found through the I.M.A.G.E. Consortium/ILNLT at:
				http://image.llnl.gov
				Plate: ILNLT0743 row: 1 column: 20
				High quality sequence atop: 651.
				Location/Qualifiers

```

SOURCE
1. . /75
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="cazon:3606"
/clone="IMAGE:4827427"
/lab_host="DH10B"
/clone_11b="N1H_MGC_97"
/notes="Organ: testis; Vector: Bluescripter (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgagc
); Oligo-dT primed using primer 5'-ttttttttttttttt-3',
size-selected for average insert size 2.2 kb and
normalized to ROF 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carrinct, in preparation). Library
constructed by M. Brownstein (NIMH/NHORI, National
Institutes of Health). Note: this is a N1H_MGC library."

BASE COUNT      191 a      212 c      193 g      179 t

ORIGIN
Query Match      37.9%;   Score 340;  DB 10;   Length 775;
Beet Local Similarity 75.5%;   Pred. No. 1.8e-87;

```


Oy 468 AATTACTGTCAGAGGAGGCTGAGTGTGCTCCACAGACACCTTGTCCAGATATCCAT 527
 Db 516 AATTACTGTCAGAGGAGGCTGAGTGTGCTCCACAGATATCCATGCTTGTACAG----- 564
 Oy 528 TGTGATCTATGAGACCCCAACAGAGATATACAGTCCCAACAGAGGAGATATGAACTCC 587
 Db 565 TTATGTCATGAGAGCCCACTCGGGGATATGAGAGCCCACTCGGAG--TACGAGAGCC 622
 Oy 588 ACCGAGAGATATGAGAGCCCAACAGAGGAGATATGAGAGCCCACTTGTGAGATATGAGAGC 647
 Db 623 CACTGAGAGATATGAGAGCCCAACAGAGGAGATATGAGAGCC--GCTGTGGATATACAGAGC 681
 Oy 648 CCGGCTG 655
 Db 682 CTCACCTG 689
 RESULT 7
 LOCUS BY098622 360 bp mRNA linear EST 07-DEC-2002
 DEFINITION BY098622 RIKEN full-length enriched, pooled tissues, adult spleen,
 etc. Mus musculus cDNA clone K630123B09 5', mRNA sequence.
 ACCESSION BY098622
 VERSION BY098622.1 GI:26209239
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 360)
 Okazaki, Y., Furuno, M., Saekawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, T., Kiyosawa, H.,
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
 Gotohori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A.,
 Gough, J., Gilmour, S., Gustinich, S., Hirokawa, N., Jackson, I.J.,
 Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M.,
 King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,
 P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki,
 H., Nagashima, T., Numata, K., Okido, T., Pavan, M.J., Pettes, G.,
 Pesole, G., Petrovsky, N., Pillai, R., Ponting, J.U., Qi, D.,
 Rameshchandra, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring,
 B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou,
 M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale,
 R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y.,
 Watanabe, Y., Wells, C., Wilming, L.G., Wyszynski, B., Yanagisawa,
 M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,
 Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Komno, H., Nakamura,
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Aikawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,
 K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,
 E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 JOURNAL MEDLINE
 PUBMED 12466851
 COMMENT
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Aizawa, K., Aikawa, T., Aikawa, T., Carninci, P., Fukuda, S., Hirozane,
 T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Komno, H., Miyazaki, A.,

Mureta, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,
 Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,
 M., Waki, K., Watabiki, A., Watanabe, M. and Hayashizaki, Y. Direct
 Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

FEATURES
 source
 location/qualifiers
 1..360
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="K630123B09"
 /clone_1b="RIKEN full-length enriched, pooled tissues,
 adult spleen, etc."
 /note="(dev stage=adult, tissue type=spleen, sex=male),
 (dev stage=adult, tissue type=kidney, sex=male),
 (dev stage=adult, tissue type=testis, sex=male),
 (dev stage=adult, tissue type=thymus, sex=male),
 (dev stage=adult, tissue type=heart, sex=male),
 (dev stage=adult, tissue type=colon, sex=male),
 (dev stage=adult, tissue type=stomach, sex=male),
 (dev stage=adult, tissue type=liver, sex=male),
 (dev stage=13 days embryo, tissue type=whole body, sex=mix
), (dev stage=16 days embryo, tissue type=whole body, sex=mix
), (dev stage=17 days embryo, tissue type=whole body, sex=mix
), (dev stage=15 days pregnant, adult, tissue type=emtion
), (dev stage=15 days pregnant, adult, tissue type=emtion
), (dev stage=10 days neonate, tissue type=brain
 , sex=mix), (dev stage=10 days neonate, tissue type=thymus
 , sex=mix), (dev stage=10 days neonate, tissue type=heart
 , sex=mix)"
 BASE COUNT 89 a 91 c 94 g 86 t
 ORIGIN
 Query Match 21.1%; Score 189.8; DB 13; Length 360;
 Best Local Similarity 76.4%; Pred. No. 5.5e-44;
 Matches 233; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 Oy 1 ATGGCAATGAAACCAAGCAACGAGGCGCTGTGGGGCCCTCATCCCTCTGGGAA 60
 Db 56 ATGGCAATGAAACCAAGCAACGAGGCGCTGTGGGGCCCTCATCCCTCTGGGAA 115
 Oy 61 AGTGTCTTGAAGAGTGTGAGATGTGACCTCTGCTTCTACAGAAACCAAGTGAATCC 120
 Db 116 AGTGTCTTGAAGAAATTTTCAAGAGTGAATCTCTCTTCCCAAGTACCCAGAGCTCC 175
 Oy 121 TATCTCTTAAATGCAAAAGAAAGAAAGCTTTTCTCACTTCATACCGAGTGTTC 180
 Db 176 AACTCTTTAGTGTGTAAGAAAGAGGGGGCTCTGTTTCTCACTTCATACCGAGTGTTC 235
 Oy 181 GTGACTTCACTTATGACCAATGACCCCATCTTTCTTATGATGCCGTTTGGCTATG 240
 Db 236 GTGACTTGGGTGACCAAGATCCATGTTTCTTATGATGCCATTTTCATATATG 295
 Oy 241 AGTGAATGCAATGAAACCAATTTTGGCCCACTACATTAAGAAACCAATTCAG 300
 Db 296 AATACTGCACTGTGAAACCAATCTTGTGGCAACATCACTTAAGGAACATTCAG 355

QY 301 GCAGC 305
 DB 356 GCAGC 360

RESULT 8
 BUI07273 932 bp mRNA linear EST 25-NOV-2002
 LOCUS 60311954P1 CSBQCHL12 Gallus gallus cDNA clone CHEST60k13 5', mRNA
 DEFINITION sequence.
 ACCESSION BUI07273.1 GI:25309754
 VERSION BUI07273.1
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE
 1 (bases 1 to 932)
 Boardman, P.B., Sanz-Bzquez, J., Overton, I.M., Burt, D.W., Bosch, B., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22355534
 12445392

TITLE JOURNAL
 MEDLINE
 PUBMED

COMMENT
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..932
 location/Qualifiers
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST60k13"
 /dev_stage="36"
 /lab_host="DH10B"
 /clone_1lb="CSBQCHL12"
 /note="Organ: heart; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; Modification of pBluescript II KS(+) and DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387,624. Cut pBluescript II KS(+) with NotI and EcoRI. Ligate in double stranded adaptor containing BspGI and BamHI sites (5'ggcgcgtcgagccgcgagatccgaaaaag) (3'aattctttctcgatccggcgctgcagcg)."

BASE COUNT 215 a 273 c 233 g 211 t

ORIGIN

Query Match 20.4%; Score 183.6; DB 13; Length 932;
 Best Local Similarity 59.5%; Pred. No. 6.1e-42;
 Matches 350; Conservative 0; Mismatches 229; Indels 9; Gaps 2;

QY 7 GTGAACGAGCCACGACGAGCGCTGCGGGCCCTCATCCCTCTGGCGAAAGTGC 66
 DB 21 GGGCTCAATAGGACCACTCGAAGAGGCGGTCTCTCTCCCAATGCCAGAGCAT 80
 QY 67 TTGAAGCATGTGAGATGTGACCTCTGCTCTCTAAGAAACGATGGAATCTATCTC 126
 DB 81 CTCGAAGCATGTGAGATGTGAGCTCTCTCTCAATGATGACGCGCAAGCTTGAAGCC 140
 QY 127 TTTAATGGCAAAAGAAAGAAAGTGTCTTCACTCACTCAACGGGTGGTCTTGAGACT 186
 DB 141 TTCAAAGGCACAAAGAAAGAAAGTGTGATCTCAACCCCATACAGAGTATCTTGATC- 199

QY 187 TCACATTAATGATGACCCCATGCTTTTATATGATGCGCTTGACCTGATGATGAC 246
 DB 200 -----AAAGGCAAGATCTCATGCTGCTTTATATGATGCGCTTTATTTGGTGAAGG 254
 QY 247 TGACCATTTGAACCAATTTTGGCCCAACATTAAGGAACCATTCAGGAGCT 306
 DB 255 TGCTATTTAGAGCGCTCTTTCTGCTTAATTAATCAAGGACGATTCAGGAG 314
 QY 307 CCAAGTGTGCTGGGAAGACAGCTGTTTTTAAGTTATCTTCAGAAAGAGTGCC 366
 DB 315 GCAGAGAGTGTGCGAAGGAGGAGGAAAGTTTAACTGACTTCAACAGGAGGAGCC 374
 QY 367 ATCAATTTGCCCACTGATGTTAAAGCTGCTCTGCTGCTGCAAGAGATTCATT 426
 DB 375 ATCAATTTGACAGCTGATGTTAAAGCTGCTCTGCTGCTGCAAGAGATTCATT 434
 QY 427 GAAGTGAATTAATTCGCTTCAACACTTCAGACTGATCAATTAATTCGCTCCAGGGCT 486
 DB 435 CAGCCCCCTGCTATGATGATCAACCTGTAACCTGAGAGGATGACCTGCTCACTGCT 494
 QY 487 GCATGCTGCTCTCAAGACAC--TTGTCACATATTCATTTGATCTATGACCC 543
 DB 495 CCAAGAGGATGACCTGCTCCAGAGGATGACCTGCTCCAGAGGATGACCT 554
 QY 544 CCACCAAGATATGATGCTCAACAGGAGGATATGAACTCCACCA 591
 DB 555 GCTCCAGAGAGGATGACCTGCTCCGAGGAGGATGATGCTCTCTCA 602

RESULT 9
 BUI03740 560 bp mRNA linear EST 25-NOV-2002
 LOCUS 603005979P1 CSBQCHL01 Gallus gallus cDNA clone CHEST3n20 5', mRNA
 DEFINITION sequence.
 ACCESSION BUI03740.1 GI:25304305
 VERSION BUI03740.1
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE
 1 (bases 1 to 560)
 Boardman, P.B., Sanz-Bzquez, J., Overton, I.M., Burt, D.W., Bosch, B., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22355534
 12445392

TITLE JOURNAL
 MEDLINE
 PUBMED

COMMENT
 Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..560
 location/Qualifiers
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST3n20"
 /issue_type="whole embryo"
 /dev_stage="20-21"
 /lab_host="DH10B"
 /clone_1lb="CSBQCHL01"
 /note="Organ: whole embryo; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; Modification of pBluescript II KS(+) (Stratagene) vector to accommodate cDNA produced with the T-trimmed protocol (construction of messenger RNA uni-directionally cloned cDNA libraries from messenger RNA

uni-directionally cloned cDNA libraries from messenger RNA

TITLE Gallus gallus bursa lymphocyte EST
JOURNAL Unpublished
COMMENT Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinstr. 52, 20251 Hamburg, Germany
Email: URI: <http://genetics.hpi.uni-hamburg.de/dt40est.html>
Location/Qualifiers

FEATURES
source 1..723
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="31her1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/clone_1lb="11ken1"
/note="CB indred strain"

BASE COUNT 161 a 205 c 181 g 172 t 4 others
ORIGIN

Query Match 20.0%; Score 179.4; DB 9; Length 723;
Best Local Similarity 66.1%; Pred. No. 8.8e-41;
Matches 277; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 7 GTGAACGAGCCACGACCGGCGTGTGGGCGCTTCATCCCTCTGGCGAAAGTTC 66
DB 15 GCGCTCAATAGGCACTCGAAGAGCGGTGTGTCTCCCATGCGAGACAT 74
QY 67 TTGAAGCAGTGTGAGATGTGACCTCTGCTTCCACGAAACGATGGAATCCATCTC 126
DB 75 CTCAGGCAATGTAAAGATGTGAGCTCTCTCACTGATGATGACGCGCAAGCTTAGGCC 134
QY 127 TTTAATGCAAAAGAAAGAAAGTGTCTCTCATCTTCATCCGGGTGTCTCTGTA 186
DB 135 TTCAAAGGCAACAAAGGAATGCTGTATCTCAACCCATACAGATATCTTCGTTC- 193
QY 187 TCACACTTAGTCATGACCCCATGCTTCTTTATGATGACCGCTTGGCTGTATGATGAC 246
DB 194 -----AAAGGCAAGATCTTATGCTGTCTTTATGATGACCGTTTATTTGTGAAAGG 248
QY 247 TGCACATTTGAACAACATTTTGGCCCACTACATTAAGAAACATTCAGGACGT 306
DB 249 TGCCTATTGAGCAGCTCTTCTTCTGTCTAATATCAATCAAGGACAGATTCAGGCTAG 308
QY 307 CCAGGTGTGCTGGGAGAGCAAGCTGTTTAAAGTATCCTTCAGAAAGAGATGCC 366
DB 309 GAGAGAGTGTGCTGGGAGAGGAGGAAAGCTTTAAAGTATTCATTCAGAGGAGAGCC 368
QY 367 ATCGAATTTGCCCACTGATGTAAAGCTGCTCTGCTGCGCAGAGAAATTCACCT 425
DB 369 ATCGAGTTTGAACAGCTGATGTTCAAAGCTGCTCTAGTGTTCAGATGAGATTCCTCT 427

RESULT 12
BI390687 568 bp mRNA linear EST 06-AUG-2001
LOCUS BI390687
DEFINITION Pp2n.pk001.p22 Normalized Chicken Pituitary/Hypothalamus/Pineal
Library Gallus gallus clone Pp2n.pk001.p22 5' similar to
gb|AA010951.1 (U79458) WW domain binding protein-2 [Homo sapiens],
mRNA sequence.
ACCESSION BI390687
VERSION BI390687.1 GI:15083969
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 568)
Porter T.B. and Cogburn L.A.
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
Library USDA/IFARs Animal Genome Project
JOURNAL Unpublished

COMMENT Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburnudel.edu, www.chickest.udel.edu
Location/Qualifiers

FEATURES
source 1..568
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Commercial broiler chicken"
/db_xref="taxon:9031"
/clone="p2n.pk001.p22"
/sex="Male and Female"
/tissue_type="Pituitary gland/Hypothalamus/Pineal gland"
/dev_stage="Embryonic (d12,d14,d19); post-hatch (w1,w3,w5,w7,w9)"
/lab_host="B. Coli BMDH108"
/clone_1lb="Normalized Chicken
Pituitary/Hypothalamus/Pineal Library"
/note="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue at different
ages. Single pass sequencing from 5'-end"

BASE COUNT 128 a 146 c 159 g 130 t 5 others
ORIGIN

Query Match 19.8%; Score 177.8; DB 12; Length 568;
Best Local Similarity 65.9%; Pred. No. 2.2e-40;
Matches 276; Conservative 0; Mismatches 137; Indels 6; Gaps 1;

QY 7 GTGAACGAGCCACGACCGGCGTGTGGGCGCTTCATCCCTCTGGCGAAAGTTC 66
DB 26 GCGCTCAATAGGCACTCGAAGAGCGGTGTGTCTCCCATGCGAGACAT 85
QY 67 TTGAAGCAGTGTGAGATGTGACCTCTGCTTCCACGAAACGATGGAATCCATCTC 126
DB 86 CTCAGGCAATGTAAAGATGTGAGCTCTCTCACTGATGATGACGCGCAAGCTTAGGCC 145
QY 127 TTTAATGCAAAAGAAAGAAAGTGTCTCTCATCTTCATCCGGGTGTCTCTGTA 186
DB 146 TTCAAAGGCAACAAAGGAATGCTGTATCTCAACCCATACAGATATCTTCGTTC- 204
QY 187 TCACACTTAGTCATGACCCCATGCTTCTTTATGATGACCGCTTGGCTGTATGATGAC 246
DB 205 -----AAAGGCAAGATCTTATGCTGTCTTTATGATGACCGTTTATTTGTGAAAGG 259
QY 247 TGCACATTTGAACAACATTTTGGCCCACTACATTAAGAAACATTCAGGACGT 306
DB 260 TGCCTATTGAGCAGCTCTTCTCTGTCTAATATCAAGGACAGATTCAGGCTAG 319
QY 307 CCAGGTGTGCTGGGAGAGCAAGCTGTTTAAAGTATCCTTCAGAAAGAGATGCC 366
DB 320 GAGAGAGTGTGCTGGGAGAGGAGGAAAGCTTTAAAGTATTCATTCAGAGGAGAGCC 379
QY 367 ATCGAATTTGCCCACTGATGTAAAGCTGCTCTGCTGCGCAGAGAAATTCACCT 425
DB 380 ATCGAGTTTGAACAGCTGATGTTCAAAGCTGCTCTAGTGTTCAGATGAGATTCCTCT 438

RESULT 13
BM490729 636 bp mRNA linear EST 07-FEB-2002
LOCUS BM490729
DEFINITION Pp2n.pk004.e6 Normalized Chicken Pituitary/Hypothalamus/Pineal
Library (Pp2n) Gallus gallus cDNA clone Pp2n.pk004.e6 5' similar
to gb|AA010951.1 (U79458) WW domain binding protein-2 [Homo sapiens]
1, mRNA sequence.
ACCESSION BM490729
VERSION BM490729.1 GI:18611660
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 636)	Phasianinae; Gallus.			
Porter,T.B. and Cogburn,L.A.				
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA				
library, USDA/IRAFs Animal Genome Project				
Unpublished				
Contact: Larry A. Cogburn				
University of Delaware				
Townsend Hall, Newark, DE 19717, USA				
Tel: 302-831-1335				
Fax: 302-831-2822				
Email: cogburn@udel.edu, www.chickest.udel.edu.				
Location/Qualifiers				
1..636				
/organism="Gallus gallus"				
/mol_type="mRNA"				
/strain="Commercial broiler chickens"				
/db_xref="taxon:9031"				
/clone="pgrp2n.pk004.e6"				
/sex="Male and Female"				
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"				
/dev_stage="Embryonic (d12,d14,d19) ; post-hatch (1,3,5,7,9 weeks)"				
/lab_host="R. coli BMDH10B"				
/clone_lib="Normalized Chicken				
Pituitary/Hypothalamus/Pineal library (pgrp2n)"				
/note="Vector: pCMVSPORT6; Library made from equivalent				
pools of total RNA isolated from each tissue at different				
ages. Single pass sequencing from 5'-end"				
BASE COUNT	143 a	174 c	173 g	146 t
ORIGIN				
Query Match	19.8%	Score 177.8	DB 12;	Length 636;
Best Local Similarity	65.9%	Pred. No. 2.4e-40;		
Matches 276; Conservative	0;	Mismatches 137;	Indels 6;	Gaps 1;
QY	7	GTGAACGACGACACCGAGAGCGCTGTGTGGGCGCCCTCATCCCTCTGGGAAAGTTC	66	
DB	26	GCGCTCATATGAGCACCACTCGAAGGAGAGCGGTGTCTGTCCCAATGCCAGAGCATT	85	
QY	67	TTGAAGCAGTGTGAGATGTGAGACTCTGCTTCTTACAGAAACAGTGAATCTATCTC	126	
DB	86	CTCAAGCAATGTAAAGATGTGAGCTCTCTTCACTGACATGACCGGACGACTTGAAGCC	145	
QY	127	TTTAATGCGACAAAGAAAGAAAGCCTTGTCTTCTCATCTTATACCGGATGTCTGTGACT	186	
DB	146	TTCAAGGACACCAAGAAAGGAAATGCTGTATCTCACCCCATACAGATATCTTGTGTCTC	204	
QY	187	TCACACTTATGATATGATACCCCATGCTTCTTTTATGATGCGGTTTGGCTGATATGATAC	246	
DB	205	-----AAGGGCAAGATCTTATGCTGTCTTTATGATGCAATTTATTTGTGAAGGGG	259	
QY	247	TGCACATTTGAACCACTATTTTGGCCGCCCACTATTAAGAAACATTCAGGACGCT	306	
DB	260	TGCTCTATTTGAGCAGCCCTTTTCTCTCTATTTACATCAAGGACAGATTCAGGCTGAG	319	
QY	307	CCAGGTGTGTGCTGGGAGAGCAAGCTGTTTTTAAGTATCTTCAAGAAAGAGTGC	366	
DB	320	GCAAGAGTGTGCTGTGAAGGGGACGGAACGTTTAAACTGATCTTTCAACAGCGGAGAGCC	379	
QY	367	ATCGAATTTGCCCACTGATGTGTAAAGCTGCTCTGTGTGTGCGACAGAAATTCACCT	425	
DB	380	ATCGAGTTTGGACAGCTGATGTTCAAAGCTGCTTATGATGCTTCAAGTGAAGTTCCTCT	438	
RESULT 14				
LOCUS	AJ454950	772 bp	mRNA	linear
DEFINITION	AJ454950 riken1 gallus gallus cDNA clone 4c7rl, mRNA sequence.			EST 22-APR-2002
ACCESSION	AJ454950			
VERSION	AJ454950.1	GI:20265046		
KEYWORDS	EST.			
SOURCE	Gallus gallus (chicken)			

ORGANISM Gallus gallus
Bakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 772)
Buerstedde, J.M.
Gallus gallus bursal lymphocyte EST
Unpublished
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinstr. 52, 20251 Hamburg, Germany
Email: URL: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.
Location/Qualifiers
1..772
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="4c7x1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/clone_id="riken1"
/note="CB inbred strain"

BASE COUNT 176 a 154 c 210 g 232 t

ORIGIN

Query Match 18.6%; Score 167.4; DB 9; Length 772;
Best Local Similarity 65.1%; Pred. No. 2.8e-37;
Matches 265; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 7 GTGAACCAAGCCACACCGAAGCCGTCGTGGGGCCCTCATCCCTCTGGCGAAAGTTC 66
DB 23 GCGCTCAATAGGCAACCACTGAAAGAAAGCGGTGTCCTGCCCAATGCCAGAGCAAT 82
QY 67 TTGAAGCAGTGTGAGATGTGACCTCTGCTTCTTACAGAAACAGTGAATCTATCTC 126
DB 83 CTCAAGCAATTAAGATGTGTGAGCTCTCTCTCAAGTCAATAGACGGCAAGCTTAGAGCC 142
QY 127 TTTAATGCACAAAAGAAAGAAACGTTTCTCATCTTATACCGAGTGTCTTCTGACT 186
DB 143 TTCAAGGCAACAAAGAAAGGAATGCTGTATCTACACCCCAATACAGAAAGATCTTGTC- 201
QY 187 TCACACTTAATGCAATGACCCCAATGCTTTTATATATGATGCCGTTTGCCCTGATGATGAC 246
DB 202 -----AAAGGCAAGATCTATGCTCTCTTTATATATGATGCCGTTTATGTAAGAGG 256
QY 247 TGCAACATTTGAACACCAATTTTGGCCCAACATTAAGAAACCATTTAGGCAAGCT 306
DB 257 TGCTCTATATGACAGCCCTCTTTCTCTGCTAATTAACAAAGACAGATTCAGGCTAG 316
QY 307 CCAGGTGTGCTGTGGAAGGACAAAGCTGTTTAAAGTTATCTTCAAGAAAGAGAGTGC 366
DB 317 GCAGAGGTGCTGTGGAAGGCAAGGCAAGCAATTTAACTTGAATTTCAACAGCGAGAGGCC 376
QY 367 ATCGAATTTGCCAAGTATGTAAGAAAGTGCCTCTGCTGCTGCGAG 413
DB 377 ATCGAATTTGACAGCTGATGTTCAAGCTGCTCTAGTGTGAAG 423

RESULT 15
LOCUS BU433499 706 bp mRNA linear EST 29-NOV-2002
DEFINITION BU433499.1 CSGBRN10 Gallus gallus cDNA clone CHS87239b14 5', mRNA
sequence.
ACCESSION BU433499
VERSION BU433499.1 GI:25922810
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Bakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 706)

AUTHORS Boardman, P. B., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E.,
Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 2235534
PubMed 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source location/Qualifiers

1..706
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHS7239b14"
/sex="Male and female"
/tissue_type="Chondrocytes isolated from growth plate cartilage"
/dev_stage="adult"
/lab_host="DH10B"
/clone_1lb="CSRB08N10"
/notes="Vector: pBluescript II KS(+), Site 1: EcoRI, Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT 189 a 140 c 179 g 198 t
ORIGIN

Query Match 18.6%; Score 166.8; DB 13; Length 706;
Best Local Similarity 65.5%; Pred. No. 4e-37;
Matches 262; Conservative 0; Mismatches 132; Indels 6; Gaps 1;

```

QY 7 GTGAACCAAGCCACACCGAGAGCCGTCGCGGAGCCCTCATCCCTCTGGCGAAAGTGTG 66
DB 27 GCGCTCAATGAGCAACATCGAAGAGGCGGTGTCTGTCTCCCAATGCCAGAGCATTT 86
QY 67 TTGAAGCAGTGTGAGATGTGACCTCTGCTTCTACAGAAACCAAGTGAATCTATCTC 126
DB 87 CTCAGCAATGTAAAGATGTGAGCTCTCTTCAAGTACATGACGGCAAGCTTGAAGCC 146
QY 127 TTTAATGCAAAAGAAAGAAAGTGTTCATCTTCACTTACCCGGTGTCTTCGAGCT 186
DB 147 TTCAAAGGACCAAGAGGAAAGTGTATCTCACCCCAATACAGATGATCTTCGTCTC- 205
QY 187 TCACACTTAGTCAATGATACCCCATGCTTTCTTTATGATGCGGTTTGACCTGATGAG 246
DB 206 -----AAAGGCAAGATCTATGCTGTCTTTTATGATGCGGTTTATTTGGTAAAGG 260
QY 247 TGCACCATGAAACCAATTTTGGCCCACTAATTAAGAAACCAATTCAGGAGCT 306
DB 261 TGCTTATTGAGAGAGCTCTTTCTCTGTAATTAATCAAGAGATTCAGGCTGAG 320
QY 307 CCAGTGTGTGCTGGGAGAGCAAGCTTTTAAAGTATCTTCAGAAAGAGAGTGC 366
DB 321 GCAGAGGTGTGCTGGGAGAGGAGCAAGCTTTAACTGACTTTCAACAGCGGAGAGCC 380
QY 367 ATCGAATTGCCCACTGATGTATAAAGTGCCTGTCTG 406

```

DB 381 ATCGAATTGAGACAGCTGATGTATAAAGCTCTCTGATG 420

Search completed: December 16, 2003, 17:31:39
Job time: 1850.09 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_p1ue_p2n model

Run on: December 16, 2003, 15:10:30 / Search time 3348.76 Seconds

(without alignments)
3823.721 Million cell updates/sec

Title: US-09-864-291-5
Perfect score: 1709

Sequence: 1 MAVVQSHTEBRGALIPSGR.....ASHRSMTAQOETSLPTSSS 313

Scoring table:

BL0SUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1p
-O=/cgn2_1/USPRO.spool/p/US09864291/runat.15122003.160858.23305/app.query.fasta_1.846
-DB=Genemb1 -QPMT=fastcap -SUFFIX=exe -MINMATCH=0.1 -LOOPCTL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.ccd1 -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09864291.QCEN_1.1354.qrunat.15122003.160858.23305 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEDBRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Genemb1:*
1: gb Da:*
2: gb Hcg:*
3: gb In:*
4: gb Om:*
5: gb Ov:*
6: gb Pat:*
7: gb Ph:*
8: gb Pl:*
9: gb Pr:*
10: gb Ro:*
11: gb Sts:*
12: gb Sy:*
13: gb Un:*
14: gb Vt:*
15: em Ba:*
16: em Fun:*
17: em Hum:*
18: em In:*
19: em Mu:*
20: em Om:*
21: em Or:*
22: em Ov:*
23: em Pat:*
24: em Ph:*
25: em Pl:*
26: em Ro:*
27: em Sts:*
28: em Un:*

29: em Vt:*
30: em Hcg_hum:*
31: em Hcg_in:*
32: em Hcg_other:*
33: em Hcg_mus:*
34: em Hcg_pln:*
35: em Hcg_rod:*
36: em Hcg_mam:*
37: em Hcg_vrt:*
38: em By:*
39: em Htgo_hum:*
40: em Htgo_mus:*
41: em Htgo_other:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1709	100.0	1413	6 AX359658	AX359658 Sequence
2	1036	60.6	2267	9 BC022546	BC022546 Homo sapi
3	1025	60.0	2266	9 BC022549	BC022549 Homo sapi
4	769	45.0	1001	6 AX359665	AX359665 Sequence
5	503	29.4	179222	2 BX296515	BX296515 Sus scrofa
6	503	29.4	222469	2 BX470149	BX470149 Danio rerio
7	475	27.8	1691	9 HS079458	HS079458 Human YW do
8	475	27.8	1820	9 BC007452	BC007452 Homo sapi
9	475	27.8	1835	9 AK057881	AK057881 Homo sapi
10	475	27.8	1912	9 BC010616	BC010616 Homo sapi
11	460	26.9	793	10 AF499026	AF499026 Rattus no
12	459.5	26.9	1885	6 AR220846	AR220846 Sequence
13	456	26.7	1802	10 MMT40826	MMT40826 Mus muscu
14	451	26.4	220895	9 HS250D1	HS250D1
15	445	26.0	2170	9 AB072784	AB072784 Macaca fa
16	424	24.8	222469	2 BX470149	BX470149 Danio rer
17	408.5	23.9	179624	3 BX296540	BX296540 Sus scrofa
18	408	23.9	19474	3 AY060828	AY060828 Drosophila
19	405	23.7	1205	5 BC053232	BC053232 Danio rer
20	392.5	23.0	168425	2 AC113593	AC113593 Mus muscu
21	392.5	23.0	224086	10 AC104325	AC104325 Mus muscu
22	377	22.1	253149	2 AC107527	AC107527 Rattus no
23	377	22.1	270171	2 AC132969	AC132969 Rattus no
24	332.5	19.5	110000	2 AC118467_0	AC118467 Mus muscu
25	321	18.8	454	6 BD027759	BD027759 Sequence
26	310	18.1	42845	2 AC017247	AC017247 Drosophila
27	310	18.1	175679	3 AC010025	AC010025 Drosophila
28	310	18.1	175590	2 AC006563	AC006563 Drosophila
29	310	18.1	297381	3 AB003540	AB003540 Drosophila
30	284.5	16.6	170781	2 AC092212	AC092212 Trypanoso
31	279	16.3	163542	2 AP001190	AP001190 Homo sapi
32	279	16.3	184371	9 AC104961	AC104961 Homo sapi
33	258	15.1	39782	2 AC138635	AC138635 Homo sapi
34	258	15.1	40476	2 AC140538	AC140538 Homo sapi
35	256.5	15.0	33899	3 CEY63D3A	CEY63D3A
36	256	15.0	148851	9 HS155D22	HS155D22
37	254.5	14.9	195859	14 AP281817	AP281817 Thupaia he
38	249	14.6	159064	9 AC138512	AC138512 Homo sapi
39	249	14.6	174921	2 AC012266	AC012266 Homo sapi
40	245.5	14.4	130742	2 AC138601	AC138601 Mus muscu
41	245.5	14.4	187391	2 AC116738	AC116738 Mus muscu
42	244.5	14.3	148055	2 AC128178	AC128178 Rattus no
43	244.5	14.3	275392	2 AC095297	AC095297 Rattus no
44	243	14.2	72300	2 AC116403	AC116403 Mus muscu
45	243	14.2	167067	10 AC124489	AC124489 Mus muscu

RESULT 1

ALIGNMENTS

AX359658
 LOCUS AX359658 1413 bp DNA linear PAT 13-FEB-2002
 DEFINITION Sequence 4 from Patent W00190185.
 ACCSSION AX359658
 VERSION AX359658.1 GI:18675409
 KEYWORDS
 SOURCE
 ORGANISM
 unclassified
 unclassified
 unclassified
 unclassified

REFERENCE
 1
 Oko, R. and Sutovsky, P.
 P132 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses thereof
 Patent: WO 0190185-A 4 29-NOV-2001;
 QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES UNIVERSITY (US)

FEATURES
 source
 1. 1413
 /organism="unclassified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 /note="mammalia"
 36. 977
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAD23336.1"
 /db_xref="GI:18675410"
 /translation="MAVNSHTSRRLGLIPSGRSLVKOCEDVDLCPLKPKVSYLFLN
 CTXKGTPLTSTRYRVPTVSHLVNDPMKSPMPRLMSDCTTBQIIPAPNTIKQITQIA
 PGGMBGQAVKLSFRKQALIEFADLMKASALAKGIPLSVNTWDTSLYITTP
 GAAVCSGQPCPAPYIVIVPPPGYTVQPSYGTIPPGYAGQKSGYGAAPMGYGAAP
 VGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGV
 NEALPPAYEASAGNTAASHRSMTAQSTSLPTSSS"

CDS
 36. 977
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAD23336.1"
 /db_xref="GI:18675410"
 /translation="MAVNSHTSRRLGLIPSGRSLVKOCEDVDLCPLKPKVSYLFLN
 CTXKGTPLTSTRYRVPTVSHLVNDPMKSPMPRLMSDCTTBQIIPAPNTIKQITQIA
 PGGMBGQAVKLSFRKQALIEFADLMKASALAKGIPLSVNTWDTSLYITTP
 GAAVCSGQPCPAPYIVIVPPPGYTVQPSYGTIPPGYAGQKSGYGAAPMGYGAAP
 VGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGV
 NEALPPAYEASAGNTAASHRSMTAQSTSLPTSSS"

BASE COUNT 377 a 363 c 369 g 304 t

ORIGIN

Alignment Scores:
 Pred. No.: 2e-79 Length: 1413
 Score: 1709.00 Matches: 313
 Percent Similarity: 100.00% Conservatve: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0

US-09-864-291-5 (1-313) x AX359658 (1-1413)

QY 1 MetAlaValaIngnInSerHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
 DB 36 ATGGCAGTGAACCGAGGCCACCGAGGCCGTCGTGGGGCCCTCATCCCTCTGGCGAA 95
 QY 21 SerValleuLysGlnCysGluAspValAspLeuGlyPheLeuGlnLysProValGluSer 40
 DB 96 AGTGTCTGAAAGCATGTGAGATGTGACCTCTGCTTCTACAGAAACCGATGGAATCC 155
 QY 41 TyrlleuPheanglyThrllylylylylylylylylylylylylylylylylylylyly 60
 DB 156 TATCTCTTAATGGCAAAAGAAAGAAAGCTTTTTCATCTTACACCGGGTGGTCTTC 215
 QY 61 ValThrSerHisLeuValaAspAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 DB 216 GTGACTTCACACTTAAGTCAATGACCCCAATGCTTCTTTAATGATGCCCTTGGCCATAG 275
 QY 81 SerAspCysThrIleGluGlnProIlePheAspProAsnTyrllelylylylylylyly 100
 DB 276 AGTGACTGACACTTGAACCAACCAATTTTGGCCCACTACATTAAAGAAACATTGAG 335
 QY 101 AlaAlaProGlyGlyGlyTyrlleGlyGlyGlyAlaAlaPheLeuSerPheArglylyly 120
 DB 336 GCAAGTCCAGAGTGGTGGTGGGAGAGCAAGCTGTTTAAATATCTTCAGAGAAAGCA 395
 QY 121 GlyAlaIleGluPheAlaGlnLeuMetValylsAlaAlaSerAlaAlaAlaArglylyle 140
 DB 396 GTGTCCATCGAATTTGCCCAACTGATGTGTAAAGCTGCTCTGCTGCGCCAGAGAAATTT 455

QY 141 ProleuGlySerValaenTyrlPheAspThrSerGlyLeuTyrlleIleThrValPro 160
 DB 456 CCACTTGAAAGTGAATTAATCTGCTTGAACATTCAGAGCACTGATACATTAATTAATCTCA 515
 QY 161 GlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrlleValIleTyrlly 180
 DB 516 GGGGCTGACAGTGTCTCTCAAGACACTTGTCCAGCATATCAATGATGATCTATGGA 575
 QY 181 ProProProProGlyTyrlThrValGlnProGlyGlyTyrllylylylylylylylyly 200
 DB 576 CCCCACACACAGATATATACATCAACAGGGGAGATATGAACTCCACAGAAAGATAT 635
 QY 201 GlyAlaGlnProGlyGlyTyrllylylylylylylylylylylylylylylylylyly 220
 DB 636 GAGGCCCAACAGGGGAGATATGAGGCCCACTATGGAATATGAGGCCGCCGCTGTGGA 695
 QY 221 Tyrllylylylylylylylylylylylylylylylylylylylylylylylylyly 240
 DB 696 TATGAGTCCCACTGAGGAGATATGAGTCCCACTGAGGAGATATGAGTCCCACTGAG 755
 QY 241 GlyTyrllylylylylylylylylylylylylylylylylylylylylylylylyly 260
 DB 756 GATATGAGGCCCACTGAGGAGATATGAGTCCCACTGAGGAGATATGAGTCCCACT 815
 QY 261 Glylylylylylylylylylylylylylylylylylylylylylylylylylyly 280
 DB 816 GGGGAGATATGAGGCCCACTGAGGAGATATGAGGCCCACTGAGGAGATATGAGGCCCTA 875
 QY 281 ProProAlaTyrllylylylylylylylylylylylylylylylylylylylylyly 300
 DB 876 CCCCCTGCAATGAAAGTCCATCTGCTGGAAGAAACAGCTCCTCAATCATATGACA 935
 QY 301 AlaGlnGlnGlnThrSerLeuProThrThrSerSerSer 313
 DB 936 GCTCAGCAGAGACTTCTCTTCCACTCACTCATCTTCT 974

RESULT 2
 BC022546 2267 bp mRNA linear PRI 04-FEB-2002
 LOCUS
 DEFINITION Homo sapiens, similar to RIKEN cDNA 4930521J23 gene, clone
 MGC:26816 IMAGE:4811804, mRNA, complete cds.
 ACCESSION BC022546
 VERSION BC022546.1 GI:18490706
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2267)
 Strausberg, R.
 Submitted (01-FEB-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: gcgaps-remail.nih.gov
 Tissue Procurement: Miklos Palokvite, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
 Tohkiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LANT)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-sngc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LANT at: <http://image.lnl.gov>
 Series: IRAX Plate: 32 Row: J Column: 20

This clone was selected for full length sequencing because it passed the following selection criteria: similarity but not identity to protein.

FEATURES

source

Location/Qualifiers

1..2267
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26816 IMAGE:481804"
/tissue_type="Brain, hippocampus"
/clone_id="N1H_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
32..961
/codon_start=1
/product="Similar to RIKEN cDNA 4930521123 gene"
/protein_id="AAH22546.1"
/db_xref="GI:18490707"
/translation="MAVNSHTNRBGLIPNGRSILKSPNVEISFPQSRGSSNVFS
GRKRTPLFTRSVRPIITSGSISDPLSPMPFDLTMTNLVEQPVZPAAPFKITQAA
PTGWSGQATPKLVFRNGALIERAQLMVKASAAAGPPLRTLNDRPSSMGIYVIG
GNNCTPOMCSYIVGAPAGYGAAPPGYGAAPGAGAPGAGAPVGNBGPVGRVRS
APLGYGAPAGYGAAPLGYGAPLGYGAPLGYGAPLGYGAPLGYGAPLGYGAPLGYG
PAGSGARPHBSTAOLPKNBASLPSASSSQVHS"

CDS

BASE COUNT 709 a 477 c 469 g 612 t
ORIGIN

Alignment Scores:

Pred. No.:	8,57e-45	Length:	2267
Score:	1036.00	Matches:	200
Percent Similarity:	73.10%	Conservative:	31
Best Local Similarity:	63.29%	Mismatches:	71
Query Match:	60.62%	Indels:	14
		Gaps:	4

US-09-864-291-5 (1-313) x BC022546 (1-2267)

QY 1 MetAlaValAlaGlnSerHisThrGluSerArgArgIValAlaLeuIleProSerGlyGlu 20
DB 32 ATGCGCGTGAATGAGAGCCACACCGAGACCGCGCGAGCCCTCATCCCTTAACGTGAA 91
QY 21 SerValLeuLeuGlyGlnGluSerValAlaLeuGluSerPheLeuGlnGluSerProValGluSer 40
DB 92 AGTCTCTTAAGAGCGCTCCGAATGAGAGCTCTCTCTCCCAAGCATCAGAGAGCTCA 151
QY 41 TyrLeuPheLeuGlnGlyGlnGluSerValAlaLeuGluSerPheLeuGlnGluSerProValGluSer 60
DB 152 AATGCTCTTAAGAGCGCTCCGAATGAGAGCTCTCTCTCCCAAGCATCAGAGAGCTCA 211
QY 61 ValThrSerHisLeuValAlaGlnSerHisThrGluSerArgArgIValAlaLeuIleProSerGlyGlu 80
DB 212 ATAACTTCATGCTCATGCTCATGCTCATGCTCATGCTCATGCTCATGCTCATGCTCATG 271
QY 81 SerAspCysThrIleGluGlnProIlePheAlaProLeuGlnGluSerPheLeuGlnGluSerProValGluSer 100
DB 272 ACCAACCCTCACTCTGAAACCAAGATTTGCTGCAAACTCTTAAGAGAGACTATTCAG 331
QY 101 AlaAlaProGlyGlyGlnGluGlnGluGlnGluGlnGluGlnGluGlnGluGlnGluGlnGluGln 120
DB 332 GCAAGCTTCATGCTCATGCTCATGCTCATGCTCATGCTCATGCTCATGCTCATGCTCATG 391
QY 121 GValAlaIleGluPheAlaGlnLeuMetValIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 140
DB 392 GCGGCCATTTGATTTGCCGATTTGATTTGCCGATTTGATTTGCCGATTTGATTTGCCGATTT 451
QY 141 ProLeuGlySerValLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
DB 452 CCACTTAAGAGCGCTCCGAATGAGAGCTCTCTCTCCCAAGCATCAGAGAGCTCA 511
QY 161 GValAlaIleAlaValCysSerSerGlnThrProCysProAlaIleThrProIleValIleIleThrGly 180
DB 512 GGG---AATATGAGCTCATGAGAGCTCTGT-----TCAGTTATTTGCTATGGA 559

QY 181 ProProProProGlyTyrThrValGlnProGlyGluTyrGlyThrProGluGlyTyr 200
DB 560 GCCCACTGCAAGAG-----TATGAGCCCACTCCCGATAC 598
QY 201 GValAlaGlnProGlyTyrGlyValAlaProPheMetGlyTyrGlyValAlaProValGly 220
DB 599 GGAGCCCACTGCAAGAGATATGAGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 658
QY 221 TyrGlyValProProGlyTyrGlyValAlaProProGlyTyrGlyValAlaProProGly 240
DB 659 TACAGAGCTCACTGTGAGATATGAGAGCCCACTCTGATATGAGAGCCCACTGCA 718
QY 241 GValTyrGlyValAlaProProGlyTyrGlyValAlaProProGlyTyrGlyValAlaProPro 260
DB 719 GATATGAGAGCCCACTGATATGAGAGCCCACTCTGATATGAGAGCCCACT 778
QY 261 GValGlyTyrGlyValAlaProProAlaGlyTyrGlyValAlaProProAlaGlyValAlaLeu 280
DB 779 CTGGATATGAGAGCCCACTCTCGATATGAGAGCCCACTGCAAGAGAGAGAGAGAGAGAGAG 838
QY 281 ProProAlaIleTyrGlyValAlaProSerAlaGlyValThrAlaIleSerHisArgSerThr 300
DB 839 CCGAGCGATATGAGAGCTCACTCTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 898
QY 301 AlaGln-----GlnGluThrSerLeuProThrThrSerSer 313
DB 899 GCCCAAGCTCTGAAAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 946

RESULT 3
LOCUS BC022549
DEFINITION Homo sapiens, similar to RIKEN cDNA 4930521123 gene, clone
ACCESSION BC022549
VERSION BC022549.1 GI:18490710
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Strausberg, R.
TITLES Direct Subtension
JOURNAL Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NIH-MGC Project URL: <http://mgc.mol.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palakovic, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) medgen@stanford.edu
Dickson, M., Schmitz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAK Plate: 32 Row: m Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: similarity but not
identity to protein.

FEATURES

source

Location/Qualifiers

1..2266
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26816 IMAGE:481804"

CDS

```

/cisue type="Brain, hippocampus"
/clone_id="NH_MGC_95"
/lab host="DH10B"
/notes="Vector: pBluescript"
32..634
/codon_start=1
/product="Similar to RIKEN CDNA 4930521123 gene"
/protein_id="AAH22549.1"
/db_xref="GI:18490711"
/translation="MAVNSHTMRRLIPIKRSGLIKRSPVELSPRSGSNVPS
GRKTTPLTSTRTVITPITSCISIDPMLSPMPPLMTLTITRQVPAANTKTIQAA
PYGMEGQATKLVPRNGALIFPAQIMVKAASAAAGPPLTLNDWFSNGIYITK
GICLHRLVCLSMKPHIDMEPHLPTEPHLDMEBNP"

```

BASE COUNT 709 a 477 c 468 g 612 t
ORIGIN

Alignment Scores:
Pred. No.: 3,13e-44 Length: 2266
Score: 1025.00 Matches: 200
Percent Similarity: 73.10% Conservative: 31
Best Local Similarity: 63.29% Mismatches: 71
Query Match: 59.98% Indels: 15
DB: Gaps: 4

US-09-864-291-5 (1-313) x BC022549 (1-2266)

```

OY 1 MetAlaValaAngInSerHisThrGluSerArgArgIValAlaLeuIleProSerGlyGlu 20
DB 32 ATGGCGGTGATCATCAGACCAACCAAGACCGCCGAGCCCTCATTCCTTAACCGTAA 91
OY 21 SerValLeuLysGlnCysGluAspValAspLeuCysPheLeuGlnLysProValGluSer 40
DB 92 AGTCTCTTGAGCGGCTCCGAAATGTGAGGCTCTCTCCACAGCATCAGAAAGCTCA 151
OY 41 TyrLeuPheAsnGlyThrLysLysGlyThrLeuPheLeuThrSerTyrArgValAlaPhe 60
DB 152 AATGCTTAGGGTGAAGAAAGCAAGAACTGTTCCTCATCTTACCGGATGATTTTC 211
OY 61 ValThrSerHisLeuValAsnAspProMetLeuSerPheMetProPheGlyLeuMet 80
DB 212 ATTACTTCATGCTCATCATGATGATCCATGTTGCTTTTATGATGCCATTGATCTGATG 271
OY 81 SerAspCysThrIleGluGlnProIlePheAlaProAsnTyrIleLysGlyThrIleGln 100
DB 272 ACGAAGCTCATGCTTGAAACACCAAGATTTGCTGAACTCATTAAGGAACATATTCAG 331
OY 101 AlaAlaProGlyGlyGlyTTPGluGlyGlnAlaValPheLysLeuSerPheArgLysGly 120
DB 332 GCAGCTCATATGTGGCTGGAGAGAACAACTTAAATTAGTCTTCAGAAATGGA 391
OY 121 GlyAlaIleGluPheAlaGlnLeuMetValLysAlaIleSerAlaAlaArgLysIle 140
DB 392 GGTGCTATTGAATTTGGCCAGTTGATGGAAGGCTGCTGCTGCTGCCGAGGAATTT 451
OY 141 ProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleIleThrValPro 160
DB 452 CCACTTAGAAGCTTAATGACGCTTCAAGCTTCAAGGAATTAATGATTAATCACTGG-GAA 510
OY 161 GlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrProIleValIleTyrGly 180
DB 511 GGG---AATATGTGACCTCCACAGATGCTCTGT-----TCAGTTATGTCATAGGA 558
OY 181 ProProProProGlyTyrThrValGlnProGlyGlyLysTyrGlyThrProGluGlyTyr 200
DB 559 GCCCACTGCAAGG-----TATGAGCCCACTTCCGGAATTC 597
OY 201 GlyAlaGlnProGlyGlyTyrGlyAlaProPheMetGlyTyrGlyAlaProProValGly 220
DB 598 GGAAGCCCACTGCAAGGATATGAGGCCCAACCCGATGAGAAATGAAAGGCCCGCTGAGGA 657
OY 221 TyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProProGly 240
DB 658 TACAGAGCTCACTGTGCGAATATGAGGCCCACTTGTGATACGAGGCCCACTGCA 717

```

```

OY 241 GlyTyrGlyAlaProProGlyGlyTyrGlyValProProGlyGlyTyrGlyAlaProPro 260
DB 718 GATATATGAGACCCCACTTATGATATGAGACCCCACTTGTGATATGAGACCCCACTT 777
OY 261 GlyGlyTyrGlyAlaProProAlaGlyTyrGlyAlaProProAlaGlyAlaLeu 280
DB 778 CTGGATATGAGACCCCACTTCTCGATATGAGACCCCACTTCCAGAAATGAAAGCCCG 837
OY 281 ProProAlaTyrGluAlaProSerAlaGlyAsnThrAlaAlaSerHisArgSerMetThr 300
DB 838 CTGGCGGATATCAGAGCTTCACTGCTGATATCAGAGCCCTCAGAAATCTTACAGCA 897
OY 301 AlaGln-----GlnGlnThrSerLeuProThrThrSerSerSer 313
DB 898 GCCAGGCTCTCGAAACAGAGGCTTCTTCCCTCGCTGCTCTTCT 945

```

RESULT 4
AX359665
LOCUS
DEFINITION
Sequence 11 from Patent WO0190185.
ACCESSION
AX359665.1 GI:18675411
VERSION
AX359665.1 GI:18675411
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1
Oko, R. and Sutovsky, P.
PC12 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses thereof
Patent: WO 0190185-A 11 29-NOV-2001;
QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES UNIVERSITY (US)
LOCATION/Qualifiers
1..1001
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1..708
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD23337.1"
/db_xref="GI:18675412"
/translation="MPPDLATNLTVRGVPVPAANFKITQIAPYGMGQATPLVER
NGDAIRERQIMYTAASAVARGPPLRTLNMPFSMGIVITIGKMGCTPOMPSCVIVYG
APRAGTAPPPGIGAPAGIGAGPVGANGPVPGRASPVATGAPPLGTGAPRAGTAP
PLGAGAPLGTGTPPLGTGAPPLGTGAPPPANGPAPGAPRASPAGSGARPQESTAAQA
PENRASPASBSGCVHS"

FEATURES
source
CDS
1..1001
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1..708
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD23337.1"
/db_xref="GI:18675412"
/translation="MPPDLATNLTVRGVPVPAANFKITQIAPYGMGQATPLVER
NGDAIRERQIMYTAASAVARGPPLRTLNMPFSMGIVITIGKMGCTPOMPSCVIVYG
APRAGTAPPPGIGAPAGIGAGPVGANGPVPGRASPVATGAPPLGTGAPRAGTAP
PLGAGAPLGTGTPPLGTGAPPLGTGAPPPANGPAPGAPRASPAGSGARPQESTAAQA
PENRASPASBSGCVHS"

BASE COUNT 261 a 254 c 248 g 238 t
ORIGIN

Alignment Scores:

Pred. No.: 1.8e-31 Length: 1001
Score: 769.00 Matches: 148
Percent Similarity: 68.60% Conservative: 18
Best Local Similarity: 61.16% Mismatches: 62
Query Match: 45.00% Indels: 14
DB: Gaps: 4

US-09-864-291-5 (1-313) x AX359665 (1-1001)

```

OY 75 MetProPheGlyLeuMetSerAspCysThrIleGluGlnProIlePheAlaProAsnTyr 94
DB 1 ATGCCATTATATCTGATGACCAACCTCATCTTAACACCAAGATTTGCTGCAAACTTC 60
OY 95 IleLysGlyThrIleGlnAlaAlaProGlyGlyGlyTTPGluGlyGlnAlaValPheLys 114
DB 61 ATTAAGGGAACCTATTCAGGAGCTCCATATGTGGCTGGAGAGAACAACTTATTA 120
OY 115 LeuSerPheArgLysGlyGlyAlaIleGluPheAlaGlnLeuMetValLysAlaIleSer 134

```

Db 121 TTAGTCTTCAAGAAATGAGATGCAATGAAATTTGCCAGTTGATGATGAAAGCTGCTT 180
Qy 135 AAlaAlaAlaArgLyIleProLeuGlySerValAsnTyTrpPheAspThrSerGlyLeu 154
Db 181 GCTGTGGCCGAGAGATTTCCACTTGAACCTTAATGATACCTTGACCTTATGAGAAAT 240
Qy 155 TTTlelleTherValProGlyAlaAlaValCysSerSerGlnThrProCysProAlaTy 174
Db 241 TATGTATTTACTCGGAGAGG---AATATGTCATCTCCACAGATGCTTT----- 288
Qy 175 ProileValIleTyrglyProProProGlyTyTrhValGlnProGlyGlyTyrgly 194
Db 289 TCAGTATTTGTCTATGAGGAGGCCCACTGACAGA-----TATGGA 327
Qy 195 ThrProProGlyIleTyrglyAlaGlnProGlyTyTrhValAlaProProMetGlyTy 214
Db 328 GCCCACCTCCGATATGAGAGCCCACTGACAGATATGAGCCCAACCCGTAGAGAAAT 387
Qy 215 GYAlaProProValGlyTyrglyValProProGlyTyTrhValAlaProProGlyTy 234
Db 388 GAAGGCCGCTGTGGATACAGAGCTCACCTGTGATATGAGGCCCACTCTTGA 447
Qy 235 TyrglyValProProGlyTyrglyValaProProGlyTyTrhValAlaProProGly 254
Db 448 TACGAGAGCCCACTGACAGATATGAGGCCCACTCTAGATATGAGGCCCACTCTT 507
Qy 255 GYTyrglyValaProProGlyTyrglyValaProProAlaGlyTyrglyValaProPro 274
Db 508 GATATGAGAACCCCACTCTCGATATGAGGCCCACTCTCGATATGAGGCCCACTCT 567
Qy 275 AlaGlyAaGlnAlaLeuProProAlaTyrglyAlaProSerAlaGlyAaThrAla 294
Db 568 GCAGAAATATAGGCCCGCTGCGGATACAGAGCTCACCTGTGATATGAGGCCAG 627
Qy 295 SerThleArgSerMetThrAlaGln-----GlnGlnThrSerLeuProThrSer 311
Db 628 CTCGAGAACTTACAGAGGCCCACTCTCGAAAGAGAGCTCTCTCTCTGCTGCTCC 687
Qy 312 SerSer 313
Db 688 TCTTCT 693
RESULT 5
LOCUS BX296515 179222 bp DNA linear HTG 26-MAY-2003
DEFINITION Sus scrofa clone PigR-121D21, *** SEQUENCING IN PROGRESS ***, 6
unordered pieces.
ACCESSION BX296515
VERSION BX296515.6 GI:31076160
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Butheria; Cetartiodactyla; Suidae; Suidae; Sus.
1 (bases 1 to 179222)
Tracy, A.
Direct Submission
Submitted (25-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 26, 2003 this sequence version replaced gi:31043704.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BT121D21
----- Summary Statistics
Assembly program: XGAP4, version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 177502 bases at least Q40
Consensus quality: 177971 bases at least Q30

Consensus quality: 178283 bases at least Q20
Insert size: 178722; sum-of-configs
Insert size: 167315; 12.1% error; agarose-fp
Quality coverage: 7.82x in Q20 bases; sum-of-configs Quality
coverage: 8.59x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 configs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the configs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 5850: contig of 5850 bp in length
* 5851 gap of 100 bp
* 5951 74264: contig of 68314 bp in length
* 74265 74364: gap of 100 bp
* 74365 84656: contig of 10292 bp in length
* 84657 84756: gap of 100 bp
* 84757 108493: contig of 23737 bp in length
* 108494 108593: gap of 100 bp
* 108594 114454: contig of 5861 bp in length
* 114455 114554: gap of 100 bp
* 114555 179222: contig of 64668 bp in length.
FEATURES
Location/Qualifiers
1..179222
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9823"
/clone="PigR-121D21"
/clone_1fb="PigR"
1..5850
/note="assembly_fragment:02838
fragment_chain:1"
5951..74264
/note="assembly_fragment:01479
fragment_chain:1"
74265..84656
/note="assembly_fragment:01113
fragment_chain:1"
84757..108493
/note="assembly_fragment:00846
fragment_chain:1"
108594..114454
/note="assembly_fragment:01058
fragment_chain:2"
114555..179222
/note="assembly_fragment:02416
fragment_chain:2"
BASE COUNT 44892 a 44880 c 44117 g 44831 t 502 others
ORIGIN
Alignment Scores:
Pred. No.: 1.05e-15 Length: 179222
Score: 503.00 Matches: 101
Percent Similarity: 58.56% Conservative: 5
Best Local Similarity: 55.80% Mismatches: 33
Query Match: 29.43% Indels: 42
DB: 2 Gaps: 4
US-09-864-291-5 (1-313) x BX296515 (1-179222)
Qy 175 ProileValIleTyrglyProProProGlyTyTrhValGlnProGlyGlyTyrgly 194
Db 55265 CCAAGTTGTGTCTATGAGCCCTTACAGATATGAGGCCCAACCAAGATATTA 55324
Qy 195 ThrProProGlyTyrglyValaGlnProGlyTyTrhValAlaProProMetGlyTy 214
Db 55325 GCCCACCTCCGATATGAGAGCCCACTCTCGATATGAGGCCCACTCTCAGAGATAC 55384
Qy 215 GYAlaProProValGlyTyrglyValaProProGlyTyTrhValAlaProProMetGlyTy 229

COMMENT

Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
 (E-mail: cdna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,
 Fax: 81-3-5449-5416)
 NRO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction and 5'-end one pass sequencing: Institute of Medical
 Science, University of Tokyo, Laboratory of Genome Structure, Human
 Genome Center; 3'-end one pass sequencing: RAB; clone selection for
 full insert sequencing: RAB and Helix Research Institute.

FEATURES

SOURCE

Location/Qualifiers

1..1835
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CB07790"
 /tissue_type="brain"
 /clone_lib="CBR"
 /note="Cloning vector: pME18SFL3"
 BASE COUNT 385 a 604 c 455 g 391 t
 ORIGIN

Alignment Scores:

Score: 3.55e-16 Length: 1835
 Percent Similarity: 475.00 Matches: 132
 Best Local Similarity: 47.08% Conservative: 37
 Query Match: 36.77% Mismatches: 113
 Indels: 77
 Gaps: 16

US-09-864-291-5 (1-313) x AK057881 (1-1835)

1 MetAlaValAngInSerHisThrGluSerArgArgAlaValAlaLeuProSerGlyGlu 20
 29 ATGGCGCTCAAGAAATCACTCGAGGCG--GGCGAGGATGCTCAAAACAGCAG 85
 21 SerValLeuYsgInCysGluAspValAspLeuCysPheLeuGlnYsProValGluSer 40
 86 AGCATCTATGCTCTATGATCATCGTGAACCTCATTCATGACATGACATGAGAACGTC 145
 41 TyrLeuPheAngGlyThrValGlyGlyThrLeuPheLeuThrSerTyrArgValAlaPhe 60
 146 GAAGCTTCAGAAAGGACCAAGAAAGGCACTGTCTACCTTACCGGCTCAATCTT 205
 61 ValThrSerHisLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 206 CTGTCCAGGCG-----AAGATGCCATGACGCTCTTCATGATCCATTTTATCTCATG 259
 81 SerAspCysThrIleGluGlnProIlePheAlaProAsnTyrIleYsgIleGln 100
 260 AAAGACTGTGATCAACAGACCCGATTTGTGTCAACTCATCAACAGGAAACAGTGAAG 319
 101 AlaAlaProGlyGlyGlyTyrGlnGlyGlnAlaValPheYsLeuSerPheArgGlyGly 120
 320 GCGGAAGCGGAGGCTGAGGAGGCTCTCTCTTCATCAAGTTGACTTTCACGCGCAGG 379
 121 GlyAlaIleGluPheAlaGlnLeuMetValYsAlaAlaSerAlaAlaArgGly--- 139
 380 GGGCGCATGATGCTTCGACAGACGCGATGCTCCAGGTGCACTTCAGGCGGTAA 439
 140 IleProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleThrVal 159
 440 GTCCCAATGAGGAGCTTATGCTTACATGACGCGGCGCTATGCTATCCCGC 499
 160 ProGlyAlaAlaValCysSerSerGlnThrProCys---ProAlaTyrProIleValIle 178
 500 CCA-----GTCCCAATGAGATGTAACCCCTGCTCCCTGCTACCC----- 541
 179 TyrGlyProProProGlyTyrThrValGlnProGlyGlyGlyTyrGlyTyrProGly 198
 542 TATCAGCGCGCCCGCCTGAGTTC-----TATCAGAGACCCCGCATG 583
 199 GlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProPro-----MetGlyTyr 214

584 ATGACCGGGCGCATG-----GGATACGTGACAGCCCCACACCGCCTTACCTGGGCC 637
 215 GlyAlaProProVal---GlyTyrGlyValPro----- 224
 638 ATGAACTCTCGATCAAGCGCGCCGATGTCTCTTCATCTCTGACCGGACCGAAGCCAGCC 697
 225 -----Pro 225
 698 GCAGAGACGACCGCGACGCTTATCAACCCAGGCAATCTTCACAAAGTCTACATGCC 757
 226 GlyGlyTyrGlyValProProGlyGlyTyrGly----- 236
 758 ACAGACGACGCGCGCATCTCTCTTACCTTACCCAGAGATGAAGAACCGATAGGCC 817
 237 -----ValProProGlyGly---TyrGlyAlaProPro----- 246
 818 CTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 877
 247 GlyGlyTyrGlyValProProGlyGlyTyrGlyAlaProProGly-----Gly 262
 878 CTGGGCTTGGGAGAGGAGGCGCGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 937
 263 TyrGlyAlaProProAlaGlyTyrGlyAlaProProAlaGlyAlaLeuProPro 282
 938 TACCAAGAACTGACATTTGTGAGCATTTAGGCGCGCGCTCTCGAGAG--GTGCGCGC 994
 282 alaTyrGlyAlaProSerAlaGlyAlaSerHisAlaArgSerMetThr 300
 995 CAGCTTCCATGCCAGCCGAGCCGACCAAGTGTGCTGCCAGCTTCTCTCTCTCTCTCT 1049

RESULT 10
 BC010616 1912 bp mRNA linear PRI 12-JUL-2001
 LOCUS Homo sapiens, clone MGC:18269 IMAGE:4177319, mRNA, complete cds.
 DEFINITION BC010616
 ACCESSION BC010616.1 GI:14714919
 VERSION MGC.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1912)
 STRAUSBERG, R.
 Direct Submission
 Submitted (10-JUL-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
 USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: gcgaps-remail.nih.gov
 Tissue Procurement: David N. Louis, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: Villalobos, tmc.edu.
 Villalobos, D.K., Luna, R.A., Hale, S.M., Huilyk, S., Lu, X., Garcia,
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
 Muzny, D.M., Gibbs, R.A.

COMMENT

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRAT Plate: 12 Row: 9 Column: 23
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis.

FEATURES

SOURCE

Location/Qualifiers
 1..1912
 /organism="Homo sapiens"

Alignment Scores:

Pred. No.: 9.32e-16 Length: 793
 Score: 460.00 Matches: 115
 Percent Similarity: 49.68% Conservative: 41
 Best Local Similarity: 36.62% Mismatches: 86
 Query Match: 26.92% Indels: 72
 DB: 10 Gaps: 12

US-09-864-291-5 (1-313) x AF499026 (1-793)

```

Oy 1 MetAlaValAenGlnSerHisTherGlnSerArgArgGlyAlaLeuIleProSerGlyGlu 20
Db 1 ATGGCGCTCAACAAAGATCACTCAAGAGGC---GGCGAGTGAATGTCACACCACTAG 57
Oy 21 SerValLeuLysGlnCysGluAspValAspLeuCySpheLeuGlnLysProValGlnSer 40
Db 58 AGCATCTTAATGCTCATATGATCATGTGAATCACTTACATTCACACGACATGAAGATTACCA 117
Oy 41 TyrLeuPheAenGlyThrLysLysGlyThrLeuPheLeuTherSerTyrArgValAlaPhe 60
Db 118 GAGGCTTCAAGAGGACCAAGAAAGGACCGCTACCTACCTACCTACCGGCTCATCTTT 177
Oy 61 ValThrSerHisLeuValAenAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
Db 178 CTGTCCAAAGGC-----AAGGATGCCATGCGGCTCTTCATATGATGCCCTTCACTCATG 231
Oy 81 SerAspCyThrIleGlnGlnProIlePheAlaProAenTyrIleLysGlyThrIleGln 100
Db 232 AAGACTGTGAGTCAAGACGCGGCTGTGTGTCCTCACTTATTAAGGGGACACTGAAA 291
Oy 101 AlaAlaProGlyGlyGlyTyrGlnGlyGlnAlaValAlaPheLeuSerPheArgLysGly 120
Db 232 GCTGAAGCAGAGAGGTGTGGAGAGGCTGCTCTCTACACCTTACACCTTACACAGAGG 351
Oy 121 GlyAlaIleGlnPheAlaGlnLeuMetValLysAlaAlaSerAlaAlaAlaArgLys 139
Db 352 GGTGCATTTGAGTTGGGACAGAGATGCTCCAGGTGCATCTCAAGCCTCCAGAGGTGA 411
Oy 140 IleProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleIleThrVal 159
Db 412 GTCCCAATGAGACCTATGAGTACCTTACATGACCAGCGGGGCGCTATGCTTCCCCCA 471
Oy 160 ProGlyAlaAlaValCysSerSerGlnThrProCys---ProAlaTyrProIleValIle 178
Db 472 CCA-----GTCCCAATGAGATGATACCTCTGCTCCCTGAGTACCC----- 513
Oy 179 TyrGlyProProProProGlyTyrThrValGlnProGlyGlyTyrThrProProGlu 198
Db 514 TATCCACCGCGCCCGCCAGCGAGTTCC----- 537
Oy 199 GlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProProMetGlyTyrGlyValProPro 218
Db 538 -----TACCTGGGCGCTCATATGATGAGCGGGGCG----- 567
Oy 219 ValGlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValPro 238
Db 568 ATGGGATATGTAAGCGCCCA----- 588
Oy 239 ProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyVal 258
Db 589 -----CCACCA----- 594
Oy 259 ProProGlyGlyTyrGlyValProPro-----AlaGlyTyrGlyValProProAlaGly 276
Db 595 CCTTATCTGGGCGCCATGAGGCTCCCGTGTCAAGGCGCCCAAGTGCCTCCCTCTACT--- 651
Oy 277 AsnGluAlaLeuProProAlaTyrGlnAlaProSerAlaGlyAsnThrAlaAlaSerHis 296
Db 652 -----CTGCACTGAGGCGCAAGGCTGCAAGACCGGCTGCCAGCGCTAT 696
Oy 297 ArgSerMetThrAlaGlnGlnGlnThrSerLeuProThrThr 310
Db 697 TACAACCGCGGCAACCAACATGCTACATGCCACGAGC 738

```

RESULT 12
 AR220846 1885 bp DNA linear PAT 26-SBP-2002
 LOCUS Sequence 87 from patent US 6426186.
 DEFINITION AR220846
 ACCESSION AR220846
 VERSION AR220846.1 GI:23327723
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE
 1 (bases 1 to 1885)
 AUTHORS Jones,K.A., Volkmuth,W. and Walker,M.G.
 TITLE Bone remodeling genes
 JOURNAL Patent: US 6426186-A 87 30-JUL-2002;
 FEATURES
 source 1..1885
 /organism="unknown"

BASE COUNT 384 a 610 c 476 g 415 t
 ORIGIN

Alignment Scores:

Pred. No.: 2.26e-15 Length: 1885
 Score: 459.50 Matches: 129
 Percent Similarity: 45.43% Conservative: 40
 Best Local Similarity: 34.68% Mismatches: 100
 Query Match: 26.89% Indels: 104
 DB: 6 Gaps: 17

US-09-864-291-5 (1-313) x AR220846 (1-1885)

```

Oy 1 MetAlaValAenGlnSerHisTherGlnSerArgArgGlyAlaLeuIleProSerGlyGlu 20
Db 61 ATGGCGCTCAACAAAGATCACTCGAAGGC---GGCGAGTGAATGTCATTAACACGAG 117
Oy 21 SerValLeuLysGlnCysGluAspValAspLeuCySpheLeuGlnLysProValGlnSer 40
Db 118 AGCATCTTAATGCTCATATGATCATGTGAATCACTTACATTCACATGAAGAACGTGCA 177
Oy 41 TyrLeuPheAenGlyThrLysLysGlyThrLeuPheLeuTherSerTyrArgValAlaPhe 60
Db 178 GAAGCTTCAAGAGGACCAAGAAAGGACCTATCACTTACCTTACCGGCTCATCTTT 237
Oy 121 GlyAlaIleGlnPheAlaGlnLeuMetValLysAlaAlaSerAlaAlaAlaArgLys 139
Db 352 GCGAAGCGGAGGAGGTGGAGAGGCTCTCTCTTCAAGTTGAATGATCAAGGCAAGG 411
Oy 140 IleProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleIleThrVal 159
Db 472 GTCCCAATGAGACCTATGAGTACCTTACATGACCAGCGGGGCGCTATGCTATGCCCG 531
Oy 160 ProGlyAlaAlaValCysSerSerGlnThrProCys---ProAlaTyrProIleValIle 178
Db 532 CCA-----GTCCCAATGAGATGATACCTCTGCTGCTACCC----- 573
Oy 179 TyrGlyProProProPro-----Gly 185
Db 574 TATCCACCGCGCCCGCCAGGTTCTATCCAGAACCCCGCATGATGAGCGGCGCATGGA 633
Oy 186 TyrThrValGlnProGlyGlyTyrThrProProGlyGlyTyrThrProProGlyGly 205

```

Db 634 TAC---GTGACAGCC---CCACACAGCCCTAC---CTGGG 666
 Qy 206 G1YTYRGLYAlaPro-----210
 Db 667 CCATGAACTCCGTTCAAGCGGCCGAGTGTCTCCACTCTGTGAGCCGAAAGCA 726
 Qy 210 -----210
 Db 727 GAGCGAGAGAGAGCCGCGCCTATTACACCCAGCAATCTCTACAGCTTACAT 786
 Qy 211 -----Prometg1YTYRGLYAlaProProVal 219
 Db 787 GCCCAGAGAGAGCCGCGCCTCTCTACTACCCAGGAGAGATAGAAAGACCACTA 846
 Qy 220 G1YTYRGLYAlaProProG1YTYRGLYAlaProProG1YTYRGLYAlaProPro 239
 Db 847 GGC-----CCTCCGCGC-----TCCGCGCTCCACCCCTGATCTCTACCTAC 891
 Qy 240 G1YTYRGLYAlaProProG1YTYRGLYAlaProProG1YTYRGLYAlaProPro 259
 Db 892 CCTTCCATCGGGGCTGTGCTGGGCTTG-----GAGGGAGAGGGGCGCT 939
 Qy 260 -----ProG1YTYRGLYAlaProProAla 268
 Db 940 TGTCTCCCTCCAGGTGTGATCATTAACATTAACA-----GAACTAGCATGTG 990
 Qy 269 G1YTYRGLYAlaProProAlaG1YAlaProProAlaProProAlaProSer 288
 Db 991 GAGCACTTA-GGGCCCCCGCGCTCGGAGG--GTGCGCGCCAGCTCCATGCGAGCC 1046
 Qy 289 AlaGlyAenThrAlaIleSerHisArgSerMetThr 300
 Db 1047 GAGCCCAAGTGTGCGCAGCGCTACTCTCCACCC 1082
 RESULT 13
 LOCUS MMU40826 1802 bp mRNA linear ROD 04-MAR-2003
 DEFINITION Mus musculus Wm-domain binding protein 2 mRNA, complete cds.
 ACCESSION U40826
 VERSION U40826.1 GI:1777578
 KEYWORDS
 SOURCE
 ORGANISM Mus musculus (house mouse)
 Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE
 AUTHORS Sudol, M., Chen, H.-I., Bougeret, C., Elmhond, A. and Bork, P.
 TITLE Characterization of a novel protein-binding module--the Wm domain
 JOURNAL PNAS Lett. 369 (1), 67-71 (1995)
 MEDLINE 95369475
 PUBMED 7641887
 REFERENCE
 AUTHORS Chen, H.-I. and Sudol, M.
 TITLE The Wm domain of Yes-associated protein binds a proline-rich ligand
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 92 (17), 7819-7823 (1995)
 MEDLINE 95372370
 PUBMED 7644498
 REFERENCE
 AUTHORS Chen, H.-I. and Sudol, M.
 TITLE Identification and Characterization of Protein Ligands to the Wm
 Domain by Western Ligand Blotting
 JOURNAL (in) Marshak, D.R. (Ed.),
 TECHNIQUES IN PROTEIN CHEMISTRY VII: 3-12;
 Academic Press, Inc., USA (1996)
 REFERENCE
 AUTHORS Chen, H.-I. and Sudol, M.
 TITLE Direct Submission
 JOURNAL Submitted (17-NOV-1995) Henry I. Chen, Laboratory of Molecular
 Oncology, The Rockefeller University, 1230 York Avenue, New York,
 NY 10021, USA
 FEATURES
 Location/Qualifiers

source
 1. 1802
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /dev_stage="16 day embryo"
 40. 825
 /note="WBP-2; putative protein ligand to the Wm domain of
 Yes-kinase associated protein YAP"
 /codon_start=1
 /product="Wm-domain binding protein 2"
 /protein_id="AA040893.1"
 /db_xref="GI:1777579"
 /translation="MALNRHSRGGVTVANNESILMSYDVELTFDMKRVFAFG
 TKGVTVLTPRVLPISKGDAMQSPFMPYLMKCEIKQVPGANPIKGIYABAG
 GMSASVYKLTPTAGALIRGCGMTQYASASRGVNWNGAYVPMSSGAYVPPVA
 NGATPCPPIPTPPPPPTPPPPMDAGNYQVPPPPYTPGPPPPVSGSAPATPA
 AEAFAAASASATYNPENPNVMPISQPPPPYTPPEDKKTO"
 BASE COUNT 396 a 565 c 437 g 404 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3.27e-15 Length: 1802
 Score: 456.00 Matches: 123
 Percent Similarity: 49.53% Conservative: 35
 Best Local Similarity: 38.56% Mismatches: 94
 Query Match: 26.68% Indels: 67
 DB: 10 Gaps: 14
 US-09-864-291-5 (1-313) x MMU40826 (1-1802)
 Qy 1 MetAlaValAngInSerHisThrGluSerArgArgG1YAlaIleProSerG1Ylu 20
 Db 40 ATGGCGCTCAACAGATCACTCAGAGGC---GGCGAGTGTGTCAACACACAGCTGAG 96
 Qy 21 SerValIleuYgInCyG1uAspValAspIleuCySerPheIleuInuYProValG1uSer 40
 Db 97 AGCATCTCAATGCTCTATGATCATGTGAGCTTACTTCAACAGACATGAGATGTGCA 156
 Qy 41 TyrIleuPheAng1YThrIleYblyG1YThrIleuPheIleuTherSeryYraGValIlePhe 60
 Db 157 GAGGCTTCACAGGAGACCAAGAAAGGACCCCTTCACTTACTCCGTACCGGTCATCTT 216
 Qy 61 ValThrSerHisIleuValAspAspProMetIleuSerPheMetProPheG1YleuMet 80
 Db 217 CTGTCTAAGGGG-----AAGACGCCATGCACTTCTTCAATGATGCTTTTACTGATG 270
 Qy 81 SerAspCyThrIleGluInProIlePheIleProAntYrIleG1YThrIleG1n 100
 Db 271 AAGACTGTGATTAAGACAGCCGGTGTTCGTCGAACTTCAATTAAGGAATGTGAAG 330
 Qy 101 AlaIleProG1Yg1YTPG1Yg1YAlaValIlePheIleuSerPheArgIleG1Y 120
 Db 331 GCTAAGACAGAGAGGTGCTGGAGAGGCTCCGCTTCAACAGTGAACCTTCAAGAGG 390
 Qy 121 G1YAlaIleG1uPheIleG1uIleuMetValIleAlaIleSerAlaIleAlaArgIY 139
 Db 391 GAGCCCATGAGTTGGACAGAGATCTCAAGGTGCATCTCAAGCTCCAGAGGGTGA 450
 Qy 140 IleProIleuG1YSerValAsnYrTrpPheAspThrSerG1YleuYrIleIleThrVal 159
 Db 451 GTCCCAATGAGAGCTATGGTACCTTACATGCCAGGAGGAGCTATGTTTCCCCCG 510
 Qy 160 ProG1YAlaValIleCySerSerG1nThrProCy---ProAlaYrProIleValIle 178
 Db 511 CCA-----GTCCCAATGAGATGTACCCCTGCTGCTACCCC-----552
 Qy 179 TyrG1YProProProG1YrYrThrValG1nProG1Yg1YrYrYrThrProProG1u 198
 Db 553 TATCCACCAACCCCAACCCAGATC-----576
 Qy 199 G1YTYRGLYAlaInProG1Yg1YTYRGLYAlaProProMetG1YTYRGLYAlaProPro 218
 Db 577 -----TATCCAGAGACTCCCATGATGATGAGGAGCC-----606

```

OY      219 ValGlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValPro 238
      607 ATGGGCTATGATACAGCCCA-----CCA 630
OY      239 ProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyVal 258
      631 CCA-----CCCTATCTCGGGCCCAAGAGCCCTCGGTCAAGGGCCCAAGTGGC 678
OY      259 ProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGly 278
      679 CCC-----GCTACTCTGCA-----GCCAGGCCAAGGCTGCAAA 714
OY      279 AlaLeuProAlaTyrGlyValProSerAlaGlyVal-----291
      715 GCGCGTCGAGTGCCTATTAACAACCA-----GGCAACCAACAATGCTACATGCC 768
OY      292 ThrAlaAlaSerHisArgSerMetThrAlaGlnGlnGlnTherLeuProThr 309
      769 ACGAGCAGCCTCCACCAACCACTACTACCCCGAGAGACAGAAAGACCCAGT 823

RESULT 14
LOCUS   H250D10                220895 bp    DNA     linear    PRI 05-JUN-2003
DEFINITION
Human DNA sequence from clone CTA-250D10 on chromosome 22. Contains
the genes for SREBF2 (sterol regulatory element binding
transcription factor 2), NAGA (alpha-N-acetylglucosaminidase), a
gene similar to neuronal-specific sepin 3, a pseudogene similar to
AN2 (adenine nucleotide translocator 2), 2 mRNAs based on ESTs, a
genomic marker D2S1178, a CA repeat polymorphism, ESTs and a CpG
island, complete sequence.
239716
ACCESSION
239716.4 GI:4456457
VERSION
HTG: ANT2; CPG Island; D2S1178; NAGA; sepin 3; SREBF2.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 220895)
REFERENCE
1 Clark G.
AUTHORS
Direct Submission
JOURNAL
Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk; Clone requests: clonerequest@sanger.ac.uk
On Mar 21, 1999 this sequence version replaced gi:4164339.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations.
corresponding to the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
CTA-250D10 is
from the human BAC library described in U-J. Kim et al. (1996)
Genomics 34, 213-218.
VECTOR: pBAC108L
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.

```

FEATURES

source

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>. This sequence is the entire insert of clone CTA-250D10. The true left end of clone RP1-18601 is at 129979 in this sequence. The true right end of clone RP5-359016 is at 1335 in this sequence. The true right end of clone RP5-921011 is at 23458 in this sequence. The location/Qualifiers

```

1..220895
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="22"
/clone="CTA-250D10"
/clone_1lb="CT19788K-A1"
17..25
/note="2.6 copies 5 mer ATTG 26% conserved"
33..325
/note="AluX repeat: matches 1..301 of consensus"
326..338
/note="2.6 copies 5 mer TTGA 26% conserved"
667..683
/note="2.1 copies 8 mer AGAGACC 34% conserved"
740..751
/note="2.4 copies 5 mer TTTA 24% conserved"
1443..1457
/note="7.5 copies 2 mer AC 21% conserved"
1715..1997
/note="Alu0 repeat: matches 5..293 of consensus"
2103..2378
/note="Alu0 repeat: matches 2..296 of consensus"
2844..2855
/note="12.0 copies 1 mer A 24% conserved"
complement(2884..3033)
/note="MIR repeat: matches 70..235 of consensus"
complement(3040..3344)
/note="AluY repeat: matches 1..310 of consensus"
3446..3681
/note="L1MB4 repeat: matches 5256..5471 of consensus"
3746..3837
/note="L1MB4 repeat: matches 5615..5707 of consensus"
3938..4142
/note="Alu0 repeat: matches 2..305 of consensus"
4143..4441
/note="L1MB4 repeat: matches 5707..5999 of consensus"
4442..4747
/note="AluSg repeat: matches 1..292 of consensus"
4748..4858
/note="L1MB4 repeat: matches 5999..6110 of consensus"
4881..5172
/note="AluX repeat: matches 1..296 of consensus"
5418..5429
/note="2.4 copies 5 mer GCTGA 24% conserved"
complement(5769..6021)
/note="match: GSS: Em:AQ069839"
6074..6089
/note="2.0 copies 8 mer CCCGGCTG 32% conserved"
6139..6153
/note="2.5 copies 6 mer TTTTA 21% conserved"
6141..6155
/note="3.0 copies 5 mer TTAA 21% conserved"
complement(6366..6572)
/note="MIR repeat: matches 5..214 of consensus"
6839..6852
/note="2.3 copies 6 mer GGAAT 28% conserved"
7088..7114
/note="6.8 copies 4 mer TTAA 36% conserved"
7091..7109
/note="1.9 copies 10 mer ATTATTAC 38% conserved"
complement(7122..7416)
/note="AluX repeat: matches 1..300 of consensus"

```


COMMENT
(E-mail: khashtinih.go.jp, URL: <http://www.nih.go.jp/yoken/genebank/>)
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181
Lab host: TOP10

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 11:00:59 ; Search time 262.166 Seconds

(without alignments)
3222.858 Million cell updates/sec

Title: US-09-864-291-5

Perfect score: 1709

Sequence: 1 MAVVQSHTRSRGALIPSGE.....ASHRSMTAQQETSLPTSSS 313

Scoring table:

BLOSUM62
Xgapop 10.0, Ygapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODL=frame+g2n.model -DEV=x1p
-Q=/cgn2_1/USPRO_epool_p/US09864291/runat_15122003_160858_23296/app_query.fasta_1.846
-DB=N Geneseq_19Jun03 -QMT=fastcap -SUPFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LISTEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonum2 -TRANS=human40.cdd
-LIST=45 -DOCALLIGN=200 -THR.SCORE=PCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=2000000000
-USER=US09864291.QCGN_1_1_511/runat_15122003_160858_23296 -NCPU=6 -ICPU=3
-NO MMAP -LARGESUBSTR -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
N Geneseq_19Jun03:*
1: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2003.DAT:*
25: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1709	100.0	1413	24 AAS20601	DNA encoding bovin
2	769	45.0	1001	24 AAS20602	DNA encoding human
3	543	31.8	467	22 AAL01229	Human reproductive
4	543	31.8	467	23 ABL96688	Human testicular a
5	475	27.8	1915	21 AAF22363	Human secreted pro
6	475	27.8	1915	25 ABZ73614	Secreted protein-e
7	475	27.8	1915	25 ABT16841	Human secreted pro
8	475	27.8	1915	25 ABZ67208	Human secreted pro
9	459.5	26.9	1885	24 ABZ70430	Human bone remodel
10	451	26.4	220895	24 ABK84798	Human CDNA differe
11	438	25.6	436	22 ABA67850	Human foetal liver
12	438	25.6	436	22 AAK42003	Human bone marrow
13	438	25.6	436	22 AAT48070	Probe #16756 used
14	438	25.6	436	22 ABY16034	Human genome-deriv
15	429.5	25.1	7099	22 AAL04882	Human reproductive
16	429.5	25.1	7099	23 ABY97776	Human testicular a
17	408	23.9	1109	23 ABY17151	Drosophila melanog
18	347	20.3	321	22 AAL01355	Human reproductive
19	347	20.3	321	22 ABL96808	Human testicular a
20	321	18.8	454	21 AAC04014	Human secreted pro
21	320.5	18.8	409	25 ABK41819	Bovine EST associa
22	310	18.1	4323	23 ABY17150	Drosophila melanog
23	231	13.5	2614	23 ABL28430	Drosophila melanog
24	231	13.5	4954	23 ABL05187	Drosophila melanog
25	231	13.5	9516	23 ABY05186	Drosophila melanog
26	228.5	13.4	894	23 AAY70582	DNA encoding novel
27	227	13.3	1863	23 ABZ28035	Drosophila melanog
28	227	13.3	4329	23 ABZ28034	Drosophila melanog
29	221.5	13.0	5751	23 ABY28034	Drosophila melanog
30	219.5	12.8	3206	23 ABY16637	Drosophila melanog
31	219.5	12.8	5595	23 ABY16636	Drosophila melanog
32	218.5	12.8	1174	21 AAC39984	Arabidopsis thalia
33	216	12.6	2016	24 ABY21419	Aspergillus oryzae
34	215	12.6	1015	21 AAF72681	Human prostate can
35	215	12.6	1812	21 AAO06029	Sequence encoding
36	214	12.5	1774	11 ABK83773	Human CDNA differe
37	214	12.5	1774	24 ABK83773	Breast carcinoma r
38	214	12.5	2176	24 ABY94224	Drosophila melanog
39	213.5	12.5	3222	23 ABY09835	Human breast cance
40	211	12.3	758	22 AAL23504	Mycobacterium tube
41	210.5	12.3	4403765	22 AAL23504	Human breast cance
42	210.5	12.3	758	22 AAL23504	Mycobacterium tube
43	209	12.2	623	23 ABY12315	Human breast cance
44	208.5	12.2	623	23 ABY12315	Drosophila melanog
45	208.5	12.2	2623	23 ABY12314	Drosophila melanog

ALIGNMENTS

RESULT 1

AAS20601 AAS20601 standard; CDNA; 1413 BP.

AC AAS20601

DT 09-APR-2002 (first entry)

XX DNA encoding bovine perinuclear theca 32 (PT32).

XX Testicular WM domain binding protein; htmpp; perinuclear theca 32;

XX PT32; contraceptive; fertility; oocyte activation; vaccine;

XX globozoospermia; spermatogenesis; spermatozoa; tyrosine kinase; c-Yes;

XX immunoncontraceptive; bovine; gene; ss.

XX OS

XX Bos sp.

FH Key Location/Qualifiers
 FT primer_bind 30..50
 FT /tag= a
 FT /note= "Primer binding site for cDNA isolation. The sequence differs from that of the forward primer given in AAS20603."
 FT CDS 35..977
 FT /tag= b
 FT /product= "PT32"
 FT /note= "Perinuclear theca 32"
 FT protein_bind 978..1001
 FT /tag= c
 FT /note= "Primer binding site for cDNA isolation. The sequence differs from that of the reverse primer given in AAS20604."
 FT MO200190185-A2.
 XX 29-NOV-2001.
 XX 25-MAY-2001, 2001MO-CA00738.
 XX 25-MAY-2000, 2000CA-2307128.
 XX 25-MAY-2000, 2000US-206979P.
 PA (TOOH) UNIV OREGON KINGSTON.
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 PI Oks R, Sutovsky P;
 DR MPI, 2002-097644/13.
 DR P-PSDB; AAU74604.
 XX
 PT Isolated perinuclear theca 32 polypeptide that interacts with activated tyrosine kinase c-Yes, for enhancing fertility, treating/diagnosing diminished fertility and abnormal spermiogenesis and for providing contraception -
 PT
 XX
 PS Claim 10; Fig 2A-B; 103pp; English.
 XX
 CC The invention describes an isolated perinuclear theca 32 (PT32) polypeptide (1) which interacts with tyrosine kinase c-Yes. (1) is useful for: enhancing fertility in a mammal; treating globozoospermia, by expressing (1) in spermatozoa; inhibiting fertilisation, by introducing (1) or its antigenic fragment into a mammal to elicit an immune response; enhancing the ability of round spermatids to activate oocytes; treating or diagnosing diminished fertility and abnormal spermiogenesis; in providing contraceptive; identifying contraceptive and fertility-enhancing agents. The polynucleotide is useful for producing (1) by recombinant techniques, as vaccine, as diagnostic reagents, and for chromosome identification. An antibody against (1) is useful in immunological assays, in immun contraceptive methods, to identify cells expressing (1), and to purify (1) by affinity chromatography. A transgenic animal is useful as an animal model for studying human fertility and reproductive biology, and for screening compounds to identify modulators of oocyte activation. The use of (1) prevents the entry of components which are detrimental to embryonic development into the oocyte during oocyte activation with crude sperm extract and avoids the propagation of viruses such as HIV (human immunodeficiency virus) and CC SIV (simian immunodeficiency virus) carried in the sperm. This sequence CC encodes the bovine perinuclear theca 32 (PT32), described in the method of the invention.
 CC
 XX
 SO Sequence 1413 BP; 377 A; 363 C; 369 G; 304 T; 0 other;

Alignment Scores:
 Pred. No.: 1,4e-86 Length: 1413
 Score: 1709.00 Matches: 313
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 24 Gaps: 0

US-09-864-291-5 (1-313) x AAS20601 (1-1413)
 QY 1 MetAlaValAsnGlnSerHisThrGluSerArgArgGlyAlaLeuLeuProSerGlyGlu 20
 DB 36 ATGGAGAGTGAACAGAGCAGACACAGAGAGCCGTGTGGGCGCTCATATCCCTGTGGCAA 95
 QY 21 SerValLeuLygGlnCysGluAspValAspLeuCysPheLeuGlnLygProValGluSer 40
 DB 96 AGTGTCTTGAAGCAGTGTGAAGATGTGAGACCTGTCTTCTTACAGAAACAGATGAAATCC 155
 QY 41 TyrLeuPheAsnGlyThrLygLygGlyThrLeuPheLeuThrSerThrArgValAlaPhe 60
 DB 156 TATCTCTTAAAGGACAAAGAAAGAAAGAGCTGTGTTTCACCTTCAATACCGGTGTCTTC 215
 QY 61 ValThrSerHisLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 DB 216 GTGACTTCACTTACTTACTTGAATGACCCATGCTTCTTCTTAAATGATCCATTTGGCTGATG 275
 QY 81 SerAspCysThrTrpLeuGlnProIlePheAlaProAsnTyrIleLygGlyThrIleGln 100
 DB 276 AGTGACTGACACATTTGAACAACCAATTTTGGCCCACTACATTAAGAAACATTCAG 335
 QY 101 AlaAlaProGlyGlyGlyTTPGluGlyGlnAlaValPheLygLeuSerPheArgLygGly 120
 DB 336 GCAAGTCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 395
 QY 121 GlyAlaIleGluPheAlaGlnLeuMetValIleValAlaSerAlaAlaArgGlyTyrIle 140
 DB 336 GGTGCCATCGAATTTTCCCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 455
 QY 141 ProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleThrValPro 160
 DB 456 CCACTTGAGAGGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 515
 QY 161 GlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrProIleValIleTyrGly 180
 DB 516 GGGGCTGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 575
 QY 181 ProProProProGlyTyrThrValGlnProGlyGlyTyrThrProProGlyGlyTyr 200
 DB 576 CCCCACACACAGATATACAGTCCACACAGGAGATATGAGATCCACAGAGAGAT 635
 QY 201 GlyAlaGlnProGlyGlyTyrGlyValAspProSerGlyTyrGlyValAspProValGly 220
 DB 636 GAGGCCAACACAGGGGATATGAGCCCACTATGAGATATGAGCCCGCTGTGGGA 695
 QY 221 TyrGlyValProProGlyGlyTyrGlyValAspProGlyGlyTyrGlyValProProGly 240
 DB 696 TATGAGATCCCACTGTGGGATATGAGTCCCACTGTGGGATATGAGTCCCACTGTGGG 755
 QY 241 GlyTyrGlyValAspProGlyGlyTyrGlyValAspProGlyGlyTyrGlyValAspPro 260
 DB 756 GGAATATGAGCCCACTGTGGGATATGAGTCCCACTGTGGGATATGAGTCCCACTGTGGG 815
 QY 261 GlyGlyTyrGlyValAspProAlaGlyTyrGlyValAspProAlaGlyTyrGlyValAsp 280
 DB 816 GGGGATATGAGCCCACTGTGGGATATGAGTCCCACTGTGGGATATGAGTCCCACTGTGGG 875
 QY 281 ProProAlaTyrGlyValAspProSerAlaGlyAsnThrAlaLeuSerHisArgSerThr 300
 DB 876 CCCCCTGATTAAGAGCTCCATCTGTGTGAATTAAGCTGTCTTCAACAGATCTATAGCA 935
 QY 301 AlaGlnGlnGluThrSerLeuProThrThrSerSerSer 313
 DB 936 GCTCAGCAGAGAGACTTCTTCTTCCCACTACATCTTCT 974

RESULT 2
 AAS20602
 ID AAS20602 strand; cDNA, 1001 BP.
 XX
 AC AAS20602;
 XX
 DT 09-APR-2002 (first entry)

XX DNA encoding human testicular WM domain binding protein (htmbp).
 XX
 XX Testicular WM domain binding protein, MBP; perinuclear theca 32;
 XX PT32; contraceptive; fertility; oocyte activation; vaccine;
 XX globozoospermy; spermatogenesis; spermatozoa; tyrosine kinase; c-Yes;
 XX immunoreceptive; human; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX primer_bind 1..18
 XX /tag= a
 XX /note= "Binds forward primer AAS20605"
 XX CDS 1..708
 XX /tag= b
 XX /product= "htmbp"
 XX /note= "Human testicular WM domain binding protein"
 XX primer_bind complement (984..1001)
 XX /tag= c
 XX /note= "Binds reverse primer AAS20606"
 XX
 XX WO200190185-A2.
 XX
 XX PD 29-NOV-2001.
 XX PD 25-MAY-2001; 2001WO-CA00738.
 XX PR 25-MAY-2000; 2000CA-2307128.
 XX PR 25-MAY-2000; 2000US-206979P.
 XX
 XX (TOOH) UNIV QUEENS KINGSTON.
 XX (UYOR-) UNIV OREGON HEALTH SCI.
 XX
 XX Olo R, Sutovsky P;
 XX
 XX MPI, 2002-097644/13.
 XX P-PSDB; AAU74610.
 XX
 XX Isolated perinuclear theca 32 polypeptide that interacts with activated
 XX tyrosine kinase c-Yes, for enhancing fertility, treating/diagnosing
 XX diminished fertility and abnormal spermatogenesis and for providing
 XX contraception -
 XX
 XX Claim 62; Fig 4B; 103pp; English.
 XX
 XX The invention describes an isolated perinuclear theca 32 (PT32)
 XX polypeptide (I) which interacts with tyrosine kinase c-Yes. (I) is
 XX useful for: enhancing fertility in a mammal; treating globozoospermy, by
 XX expressing (I) in spermatozoa; inhibiting fertilization, by introducing
 XX (I) or its antigenic fragment into a mammal to elicit an immune
 XX response; enhancing the ability of round spermatids to activate oocytes;
 XX treating or diagnosing diminished fertility and abnormal spermatogenesis;
 XX in providing contraception; identifying contraceptive and
 XX fertility-enhancing agents. The polynucleotide is useful for producing
 XX (I) by recombinant techniques, as vaccine, as diagnostic reagents, and
 XX for chromosome identification. An antibody against (I) is useful in
 XX immunological assays, in immunoreceptive methods, to identify cells
 XX expressing (I), and to purify (I) by affinity chromatography. A
 XX transgenic animal is useful as an animal model for studying human
 XX fertility and reproductive biology, and for screening compounds to
 XX identify modulators of oocyte activation. The use of (I) prevents the
 XX entry of components which are detrimental to embryonic development into
 XX the oocyte during oocyte activation with crude sperm extract and avoids
 XX the propagation of viruses such as HIV (human immunodeficiency virus) and
 XX SIV (simian immunodeficiency virus) carried in the sperm. This sequence
 XX encodes the human testicular WM domain binding protein (htmbp), described
 XX in the method of the invention.
 XX
 XX SQ Sequence 1001 BP; 261 A; 254 C; 248 G; 238 T; 0 other;
 XX
 Alignment Scores: 1.4e-34 Length: 1001
 Pred. No.:

Score: 769.00 Matches: 148
 Percent Similarity: 68.60% Conservative: 18
 Best Local Similarity: 61.16% Mismatches: 62
 Query Match: 45.00% Indels: 14
 DB: 24 Gaps: 4
 US-09-864-291-5 (1-313) x AAS20602 (1-1001)
 QY MetProPheGlyLeuMetSerAspCysThrIleGluGlnProIlePheAlaProAsnTyr 94
 DB 1 ATGCAATTGATCTGATGACGACACCTCACTGTAAACACGATTTGCTGCAAACTTC 60
 QY 11etysgIyThriIleGlnAlaAlaProGlyGlyTyrTyrGluGlnAlaValPheIys 114
 DB 61 ATTAAAGGAACTTATTCAGGACGCTCCATATGATGCTGGAGAGACAACTTCTTAA 120
 QY 115 LeuSerPheArgIysGlyGlyAlaIleGluPheAlaGlnLeuMetValIleAlaSer 134
 DB 121 TTAGCTTCAGAAATATGAGATGCAATTTGATGATGCTGTAAGCTGCTCT 180
 QY 135 A1A1A1A1A1A1GlyIleProLeuGlySerValIleTyrTyrPheAspThrSerGlyLeu 154
 DB 181 GCTGTGCCCGAGGATTTCCACTTAAATATGACTGTTCAGCTTATGGAAT 240
 QY 155 TyrIleIleThrValProGlyAlaAlaValCysSerSerGlnThrProCysProAlaTyr 174
 DB 241 TATGTAATTAAGTGGAGAAAGG--AAATATGCACTCCACAGATCCTTGT----- 288
 QY 175 ProIleValIleTyrGlyProProProProGlyTyrThrValGlnProGlyGluTyrGly 194
 DB 289 TCATTAATTCATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 327
 QY 195 ThrProProGluGlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProProMetGlyTyr 214
 DB 328 GCCCACTCCCGATATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 387
 QY 215 GlyAlaProProValGlyTyrGlyAlaProProGlyGlyTyrGlyAlaProProGly 234
 DB 388 GAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 447
 QY 235 TyrGlyValProProGlyGlyTyrGlyAlaProProGlyGlyTyrGlyAlaProProGly 254
 DB 448 TACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 507
 QY 255 GlyTyrGlyAlaProProGlyGlyTyrGlyAlaProProAlaGlyTyrGlyAlaProPro 274
 DB 508 GGAATATGGAACCCACCTCTCGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 567
 QY 275 AlaGlyValGlnAlaLeuProProAlaTyrGlyAlaProSerAlaGlyIleThrAlaAla 294
 DB 568 GCAGAAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 627
 QY 295 SerIleAspSerMetThrAlaGln-----GlnIleThrSerIleProThrThrSer 311
 DB 628 CCGAGAAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 687
 QY 312 SerSer 313
 DB 688 TCTTCT 693
 RESULT 3
 AAL01229 standard; cDNA; 467 BP.
 AAL01229;
 21-NOV-2001 (first entry)
 Human reproductive system related antigen cDNA SEQ ID NO: 1230.
 Human; reproductive system related antigen; reproductive system disorder;
 cancer; gene therapy; ss.
 XX

OS Homo sapiens.
XX
PN W0200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001W0-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190706.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218230.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231968.
PR 12-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241321.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251188.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.

```
XX PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI, 2001-465570/50.
XX P-PSDB; AA095259.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
XX PT is used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 1; SEQ ID NO 1230; 1297bp + Sequence Listing, English.
XX
XX The present invention provides the protein and coding sequences of a
XX CC number of human reproductive system related antigens. These can be used
XX CC in the prevention and treatment of reproductive system disorders,
XX CC including cancer. The present sequence is a coding sequence of the
XX CC invention.
XX
XX SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
XX
XX Alignment Scores:
XX Pred. No.: 2,38e-22 Length: 467
XX Score: 543.00 Matches: 104
XX Percent Similarity: 69.46% Conservative: 12
XX Best Local Similarity: 62.28% Mismatches: 38
XX Query Match: 31.77% Indels: 13
XX DB: 22 Gaps: 4
XX
XX US-09-864-291-5 (1-313) x AAL01229 (1-467)
XX
XX 118 ArglyGlyGlyValaIleGluPhealaGlnLeuMetValIleAlaIleAlaIleAla 137
XX DB 3 AAAAAATGAGAGTCCATTGATTTGCCCACTGATGATGAAGCTCCCTCTGCTGCGC 62
XX
XX 138 ArgGlyIleProLeuGlySerValaIleAntyTTPheapSPTHserGlyLeuYrilelle 157
XX DB 63 CAGAGATTTCACCTTGAACCTTAATGACTGGTTCAGCTCTATGGGAATTATGTAATT 122
XX
XX 158 ThrValProGlyValaIleValaIleCysSerSerGlnThrProCysProAlaIleTyrProIleVal 177
XX DB 123 ACTGGGGAGAGG---AAATATGTCACCTCCACAGATGCTTGT-----TCAGTTATT 170
XX
XX 178 IleTyrGlyProProProProGlyTyrThrValGlnProGlyValuTyrGlyThrProPro 197
XX DB 171 GTCATAGAGCCCACTGACAGA-----TATGAGCCCACT 209
XX
XX 198 GluGlyTyrGlyValaIleProGlyGlyTyrGlyValaProProMetGlyTyrGlyAlaPro 217
XX DB 210 CCCGATACGAGAGCCCACTGACAGATATGAGCCCAACCGTAGGAATGAGAGCCG 269
XX
XX 218 ProValGlyTyrGlyValaIleProProGlyGlyTyrGlyValaProProGlyGlyTyrGlyVal 237
XX DB 270 CCTGTGGATACAGAGCCCTCAGCTGTGCAATGAGAGCCCACTCTTGATACGAGGCC 329
XX
XX 238 ProProGlyGlyTyrGlyValaIleProProGlyGlyTyrGlyValaProProGlyGlyTyrGly 257
XX DB 330 CCACTCGAGATACAGAGCCCACTCTGAGATATGAGAGCCCACTTTGAGATATG 389
XX
XX 257 ValaProProGlyGlyTyrGlyValaIleProProAlaGlyTyrGlyValaProProAlaGly 277
XX DB 390 AACCCCACTCTCGATATGAGAGCCCACTTTCGATATGAGAGCCCACTGAGAGAA 449
XX
XX 277 nGluAlaLeuProProAla 283
XX DB 450 TGAAGGC---CCGCTTCG 465
XX
XX RESULT 4
XX ID ABL96688 Standard; cDNA; 467 BP.
XX
XX ABL96688;
XX
XX 21-JUN-2002 (first entry)
XX
```

```
XX DB Human testicular antigen encoding cDNA SEQ ID NO: 356.
XX
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
XX KW reproductive system disorder; urinary system disorder; gene therapy;
XX KW cardiovascular disorder; respiratory disorder; neurological disorder;
XX KW gastrointestinal disease; infection; cytostatic; gene; ss.
XX OS Homo sapiens.
XX
XX FN WO20015317-A2.
XX
XX PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001MO-US01329.
XX
XX 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226681.
XX PR 22-AUG-2000; 2000US-0226686.
XX PR 22-AUG-2000; 2000US-0227182.
XX PR 23-AUG-2000; 2000US-0227009.
XX PR 30-AUG-2000; 2000US-0228924.
XX PR 01-SEP-2000; 2000US-0229287.
XX PR 01-SEP-2000; 2000US-0229343.
XX PR 01-SEP-2000; 2000US-0229344.
XX PR 01-SEP-2000; 2000US-0229345.
XX PR 05-SEP-2000; 2000US-0229509.
XX PR 05-SEP-2000; 2000US-0229513.
XX PR 06-SEP-2000; 2000US-0230437.
XX PR 06-SEP-2000; 2000US-0230438.
XX PR 08-SEP-2000; 2000US-0231242.
XX PR 08-SEP-2000; 2000US-0231243.
XX PR 08-SEP-2000; 2000US-0231244.
XX PR 08-SEP-2000; 2000US-0231413.
XX PR 08-SEP-2000; 2000US-0231414.
XX PR 08-SEP-2000; 2000US-0232080.
XX PR 08-SEP-2000; 2000US-0232081.
XX PR 12-SEP-2000; 2000US-0231968.
XX PR 14-SEP-2000; 2000US-0232397.
XX PR 14-SEP-2000; 2000US-0232398.
XX PR 14-SEP-2000; 2000US-0232399.
XX PR 14-SEP-2000; 2000US-0232400.
XX PR 14-SEP-2000; 2000US-0232401.
XX PR 14-SEP-2000; 2000US-0233063.
```

PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236357.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.

PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 08-DEC-2000; 2000US-0251980.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI, 2001-483232/52.
 XX
 PT Nucleic acids encoding 973 human testicular antigen polypeptides,
 XX useful for preventing, diagnosing and/or treating testicular cancer -
 XX
 PS Claim 1, SEQ ID NO 356; 766bp; English.
 CC The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer,
 CC especially testicular cancers. The present sequence is a CDNA of the
 CC invention.
 XX
 XX
 SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
 Alignment Scores:
 Pred. No.: 2,386-22 Length: 467
 Score: 543.00 Matches: 104
 Percent Similarity: 69.46% Conservative: 12
 Best Local Similarity: 62.28% Mismatches: 38
 Query Match: 31.77% Indels: 13
 DB: Gaps: 4
 US-09-864-291-5 (1-313) x ABL96688 (1-467)
 QY 118 ArgLyGlyGlyValAlaIleGluPheAlaGlnLeuMetValIleAlaIleAlaAla 137
 Db 3 AGAAATGAGAGGTCCATTGATTTGGCCAGTTGATGATGAAGGCTGCTGCTGCG 62
 QY 138 ArgGlyIleProLeuGlySerValAenryrTTPheAepThSergIyLeuTyrllelle 157
 Db 63 CGAGATTTCACCTTGAACTTAATGACTGGTTCAGCTCTATGGAATTATGTAATT 122
 QY 158 ThrValProGlyValAlaIleValCysSerSergInThProCyProAlaIyrProIleVal 177
 Db 123 ACTGGGGAAGG---AATATGTGACTCCACAGATGCTTGT-----TCAGTTATT 170
 QY 178 IleTyrglyProProProProGlyTyrtThrValGlnProGlyGluTyrglyThrProPro 197
 Db 171 GTCTATGAGGCCCACTGACAG-----TATGAGGCCCACT 209
 QY 198 GluGlyTyrglyValAlaGlnProGlyGlyTyrglyValaProPrometGlyTyrglyAlaPro 217
 Db 210 CCCGATPACGAGGCCCACTGACGATATGAGCCCAACCCCTGGAATGAAAGCCCG 269
 QY 218 ProValGlyTyrglyValaProProGlyGlyTyrglyValaProProGlyGlyTyrglyVal 237
 Db 270 CCTGTGGAATACAGAGCTCACTGTGCATATGAGGCCCACTTGTGATACGAGCC 329
 QY 238 ProProGlyGlyTyrglyValaProProGlyGlyTyrglyValaProProGlyGlyTyrgly 257
 Db 330 CCACTGACGAGATATGAGGCCCACTTGAATATGAGGCCCACTTGTGATATG 389
 QY 257 yAlaProProGlyGlyTyrglyValaProProAlaGlyTyrglyValaProProAlaGly 277
 Db 390 AACCCCACTCTGATATGAGGCCCACTTGTGATATGAGGCCCACTGACAGAAA 449
 QY 277 ngIuaIaLeuProProAla 283
 Db 450 TGAAGC---CGCCTTTCG 465

QY 263 TyrGlyAlaProPro-AlaGlyTyrGlyAlaProProAlaGlyAlaSerGlyAlaLeuProP 282
 DB 968 TACCAAGAACTACATTCATTCAGGACATTCAGGACCCCGGCTCCGAGAG---GTGCGCC 1024
 QY 282 cAlaTyrGlyAlaProSerAlaGlyAlaSerAlaGlySerMetThr 300
 DB 1025 CAGCTTCATTCAGGACCCGAGGACCAAGTGTGCGGACGATCTCCCTCACC 1079
 RESULT 6
 AB273614
 ID AB273614 standard; cDNA; 1915 BP.
 AC AB273614;
 XX
 DT 12-MAY-2003 (first entry)
 XX
 DE Secreted protein-encoding gene 334 cDNA clone HTLEML6, SEQ ID NO:344.
 XX
 KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antihaemic; vulnery; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200277013-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-US09370.
 XX
 PR 27-MAR-2001; 2001US-278650P.
 PR 12-SEP-2001; 2001US-0950082.
 PR 12-SEP-2001; 2001US-0950083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2003-040578/03.
 DR P-PSDB; ABR01280.
 XX
 PT New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia -
 XX
 PS Claim 21; Page 1332; 2474pp; English.
 XX
 CC AB273281-AB273697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP0947-ABP01363 represent the proteins they encode.
 CC AB273698-AB274687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, prohormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein-encoding cDNA clone of the invention.

BQ Sequence 1915 BP; 401 A; 613 C; 481 G; 413 T; 7 other;
 Alignment Scores:
 Pred. No.: 5, 15e-18 Length: 1915
 Score: 475.00 Matches: 112
 Percent Similarity: 47.08% Conservative: 37
 Best Local Similarity: 36.77% Mismatches: 113
 Query Match: 27.79% Indels: 77
 DB: 25 Gaps: 16
 US-09-864-291-5 (1-313) x AB273614 (1-1915)
 QY 1 MetAlaValAsnGlnSerHisThrGluSerArgArgGlyAlaLeu1LeProSerGlyGlu 20
 DB 59 ATGGCGCTCAACAGAAATCATCTCGAAGGCG---GGCGAGATGATGTGTAATACACCGAG 115
 QY 21 SerValLeuLysGlnCysGluArgValAspLeuCysPheLeuGlnLysProValGluSer 40
 DB 116 AGCATCTTAATGTCTTATGATCATCGTGAACCTCAATTCATGATGATGAAGACGTGCCA 175
 QY 41 TyrLeuPheAsnGlyThrLysLysGlyThrLeuPheLeuThrSerCysArgValAlaPhe 60
 DB 176 GAAGCTTCAAGAGACCAAGAAAGCACTGTCTACCTTACCCCTTACCGGCTCATCTT 235
 QY 61 ValThrSerHisLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 DB 236 CTGTCCAGAGGC-----AAGATGCCATGACGTCTTCATGATGATTCATTCATCAG 289
 QY 81 SerAspCysThrIleGlnGlnProIlePheAlaProAsnTyrIleLysGlyThrIleGln 100
 DB 290 AAAGACTGTGAATCAGACAGACCGCTATTTGTGTGCAACTACATCAAGGACAGTGAAG 349
 QY 101 AlaAlaProGlyGlyGlyThrGluGlyGlnAlaValPheLysLeuSerPheArgLysGly 120
 DB 350 GCGAAGCGGAGAGTGGCTGGAGAGCTCTGCTTCTTCAAGTTACCTTACCGGCTCATCTT 409
 QY 121 GlyAlaIleGluPheAlaGlnLeuMetValLysAlaAlaSerAlaAlaAlaArgGly 139
 DB 410 GGCGCATTCATGATTCAGACAGCGGATGCTCCAGGTGGATCTCAAGCTCCAGAGGTGA 469
 QY 140 IleProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleIleThrVal 159
 DB 470 GTCCCCAGTGAAGCCATAGTACTCTTACAGGCCAGCGGSGCTATCTATCTATCCCCCG 529
 QY 160 ProGlyAlaAlaValCysSerSerGlnThrProCys---ProAlaTyrProIleValIle 178
 DB 530 CCA-----GTGCGCAATGAATGATACCCCTGCTCGGCTACCC----- 571
 QY 179 TyrGlyProProProGlyTyrThrValGlnProGlyGlyTyrGlyThrProProGlu 198
 DB 572 TATCCACCGCCCGCCACCTGAGTTC-----TATCCAGAGACCCCGCATG 613
 QY 199 GlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProPro-----MetGlyTyr 214
 DB 614 ATGAGACGGGGGCATG-----GATATCGTGAAGCCCCCACAAGCCCTTACCTGGGCGCC 667
 QY 215 GlyAlaProProVal---GlyTyrGlyValPro----- 224
 DB 668 ATGAACTTCGGTCAAGCGGCGCCGATGTCCCTCCATCTGCAAGCCGAGACCAAGGCC 727
 QY 225 ----- 225
 DB 728 GCAAGAAAGACCGCCAGCGCTTATCAACCAAGGCAATCTTCACAAGCTTACATGCCC 787
 QY 226 GlyGlyTyrGlyValProProGlyGlyTyrGly----- 236
 DB 788 ACAGACCAAGCCCGCCACTCTCTACTACCAACCGAAGATGAAGAACCAAGTAGGCC 847
 QY 237 -----ValProProGlyGly---TyrGlyAlaProPro----- 246
 DB 848 CTCCTGCTCCCTGCTCCCTCCATCTCTCTACCTTACCTCCATCGGAGGCTGTG 907
 QY 247 GlyGlyTyrGlyValProProGlyGlyTyrGlyAlaProProGly-----Gly 262

Db 908 CTGGGGCTTGGGGAGGGAGGGGCGCTTGTTCCTCCCGAGCTGATCATTAACAAT 967
 Qy 263 TTTGTAAlaProPro-AlaGlyTyrGlyAlaProProAlaGlyAangluAlaLeuProPr 282
 Db 968 TACCAAGAACTAGCATTTGTGGACATTAAGGCCCCCGGCTCGGAGAG--GTGCCGCC 1024
 Qy 282 GAlATyrGluAlaProSerAlaGlyAanThraAlaSerHisArgSerMetThr 300
 Db 1025 CAGCTTCCCATGCCAGCCCGAGGCCACAGTGTGCTGCCAGCTACTCTCCCTACCC 1079
 RESULT 7
 ABT16841
 ID ABT16841 standard; DNA; 1915 BP.
 XX AC ABT16841;
 XX DT 03-APR-2003 (first entry)
 XX DE Human secreted protein gene sequence - SRQ ID No 90.
 XX Human; gene; ds; protein therapy; immediate hypersensitivity disease;
 KW allergic disorder; asthmatic disorder; gene therapy; secreted protein;
 KW hay fever; allergic conjunctivitis; allergic rhinitis;
 KW binding partner identification; chromosome identification;
 KW radiation hybrid mapping; long-range restriction mapping.
 XX OS Homo sapiens.
 XX PN WO200271188-A2.
 XX PD 03-OCT-2002.
 XX PF 26-MAR-2002; 2002MO-US09239.
 XX PR 27-MAR-2001; 2001US-278650P.
 PR 12-SEP-2001; 2001US-0950082.
 XX PR 12-SEP-2001; 2001US-0950083.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Rosen CA, Ruben SM;
 XX DR WPI; 2003-175010/17.
 PT Use of human secreted proteins and nucleic acids for preparing a
 PT diagnostic or pharmaceutical composition for diagnosing or treating
 PT allergic or asthmatic disorders, e.g. asthma, hay fever, or allergic
 PT conjunctivitis or rhinitis
 XX PS Claim 7, Page 607; 823pp; English.
 CC The invention comprises the amino acid and coding sequences of human
 CC secreted proteins. The DNA and protein sequences of the invention are
 CC useful for the diagnosis and treatment of allergic disorders, asthmatic
 CC disorders and immediate hypersensitivity diseases (e.g. hay fever,
 CC allergic conjunctivitis and allergic rhinitis). The proteins of the
 CC invention are also useful for identifying a binding partner. The nucleic
 CC acids of the invention are also useful for chromosome identification.
 CC radiation hybrid mapping or long-range restriction mapping. The present
 CC DNA sequence encodes a human secreted protein of the invention.
 SQ Sequence 1915 BP; 401 A; 613 C; 481 G; 413 T; 7 other;
 Alignment Scores:
 Pred. No.: 5,156-18 Length: 1915
 Score: 475.00 Matches: 132
 Percent Similarity: 47.08% Conservative: 37
 Best Local Similarity: 36.77% Mismatches: 113
 Query Match: 27.79% Indels: 77
 DB: 25 Gaps: 16
 US-09-864-291-5 (1-313) x ABT16841 (1-1915)

Qy 1 MetAlaValaAngInSerHisThrGlyUserArgArglyAlaLeuLeuProSerGly 20
 Db 59 ATGGCGCTCAACAAAGATCACTCGAGGCC---GGCGAGAGATCTCTAAATACACCGAG 115
 Qy 21 SerValLeuLysGlnCyseGluAspValAspLeuCysePheLeuGlnLysProValGlySer 40
 Db 116 AGCATCTTAATGTCCTTAATGATCACTGGAATCTCAATTCATCAATGACATGAGAGCTGCCA 175
 Qy 41 TyrLeuPheAngInGlyThrLysLysGlyThrLeuPheLeuThrSerTyrArgValAlaPhe 60
 Db 176 GAAGCTTCAAGAGGACCAAGAAAGGACCTGTCACTTACCTTACCTTACCTTACCTTACCTT 235
 Qy 61 ValThrSerHisLeuValaAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 Db 236 CTGTCCAAAGGC-----AAGATGCCATGACATGATCTTCAATGATCCATTTTATCTCATG 289
 Qy 81 SerAspCyseThrIleGluGlnProIlePheAlaProAsnTyrIleLysGlyThrIleGln 100
 Db 290 AAAGACTGTGATGATCAAGACACCGCTATTGTGTGCAAACTACATCAAGGAGGACGTAGAG 349
 Qy 101 AlaAlaProGlyGlyGlyTyrGlyGlyGlnAlaValPheLysLeuSerPheArgLysGly 120
 Db 350 GCGAAGCGGAGGAGGTGGCTGGGAAGGCTCTCTCTTCAAGATGACTTTCACGCGAGGG 409
 Qy 121 GlyAlaIleGluPheAlaGlnLeuMetValAlaAlaSerAlaAlaAlaArgGly--- 139
 Db 410 GGCGCATTAAGTTCGACACGAGATCTCAAGGTGATCTCAAGCTTCAAGCTTCAAGCTTCA 469
 Qy 140 IlePheLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleIleThrVal 159
 Db 470 GTCCCAAGTGAAGCTATGATCTTACTTACATGACCCAGCGGCTATGTATCTATCCCGCG 529
 Qy 160 ProGlyAlaAlaValCyseSerGlnThrProCyse---ProAlaTyrProIleValIle 178
 Db 530 CCA-----GTGCGCAATGAATGATACCTCGCTGCTCTGACTACCCC----- 571
 Qy 179 TyrGlyProProProGlyTyrThrValGlnProGlyGlyTyrGlyThrProProGly 198
 Db 572 TATCCACCGCCCACTGAGTTC-----TATCCAGAACCCCCCAATG 613
 Qy 199 GlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProPro-----MetGlyTyr 214
 Db 614 ATGACCGGAGCCAGT-----GATACGTGACGCCCCACACCGGCTTACCTTGGGCC 667
 Qy 215 GlyAlaProProVal---GlyTyrGlyValPro----- 224
 Db 668 ATGGAACCTCGGTCAAGGCGCCGAGTGTCTTCACTCTCGACCGAAGCAAGGCC 727
 Qy 225 ----- 225
 Db 728 GCAGAGACCGCCGACGCGCTATTACACCCAGGCAATCTTCAACATCTTACATGATGCC 787
 Qy 226 GlyGlyTyrGlyValProProGlyGlyTyrGly----- 236
 Db 788 ACAGACGACCGCGCCGACCTCTTACTTACCCACCGAAGATTAAGAAAGACCAAGTAGGCC 847
 Qy 237 -----ValProProGlyGly---TyrGlyAlaProPro----- 246
 Db 848 CTCGCGCTTCCGTGCTTCCACCTTCAATCTCTTACCTTACCTTCCATGCGGCGCTGTG 907
 Qy 247 GlyGlyTyrGlyValProProGlyGlyTyrGlyAlaProProGly-----Gly 262
 Db 908 CTGGGGCTTGGGGAGGGAGGGGCGCTTGTTCCTCCCGAGCTGATCATTAACAAT 967
 Qy 263 TTTGTAAlaProPro-AlaGlyTyrGlyAlaProProAlaGlyAangluAlaLeuProPr 282
 Db 968 TACCAAGAACTAGCATTTGTGGACATTAAGGCCCCCGGCTCGGAGAG--GTGCCGCC 1024
 Qy 282 GAlATyrGluAlaProSerAlaGlyAanThraAlaSerHisArgSerMetThr 300
 Db 1025 CAGCTTCCCATGCCAGCCCGAGGCCACAGTGTGCTGCCAGCTACTCTCCCTACCC 1079

RESULT 8
 AB267208 standard; cDNA; 1915 BP.
 ID AB267208 standard; cDNA; 1915 BP.
 XX
 AC AB267208;
 DT 26-MAR-2003 (first entry)
 XX
 DE Human secreted protein encoding cDNA SEQ ID NO 328.
 XX
 KW Human; secreted protein; nontropic; neuroprotective; cytostatic;
 KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
 KW viral; antiparasitic; antiparkinsonian; antiobesity; antianemic;
 KW antidiabetic; cancer; antineoplastic; hepatotropic; cerebroprotective;
 KW antihypertensive; antiallergic; antidiabetic; antitumor; anticonvulsant;
 KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;
 KW cardiovascular disorder; neurological disease; nephrotropic;
 KW gene therapy; gene; de.
 XX
 OS Homo sapiens.
 XX
 PN WO200277186-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-US09188.
 XX
 PR 27-MAR-2001; 2001US-278650P.
 PR 12-SEP-2001; 2001US-0950082.
 PR 12-SEP-2001; 2001US-0950083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2003-040583/03.
 DR P-PSDB; ABP99787.
 XX
 PT New human secreted proteins encoded by genes contained in cDNA clones
 PT (e.g. HGCA13), useful for preventing, treating or diagnosing e.g.
 PT AIDS, multiple sclerosis, herpes virus, leukemia, tick-borne
 PT encephalitis or West Nile fever
 XX
 PS Claim 7; Page 1339; 2423pp; English.
 XX
 CC The invention relates to novel human genes (AB266891-AB268209) and the
 CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections.
 XX
 SQ Sequence 1915 BP; 401 A; 613 C; 481 G; 413 T; 7 other;
 Alignment Scores:
 Pred. No.: 5.15e-18 Length: 1915
 Score: 475.00 Matches: 132
 Percent Similarity: 47.08% Conservative: 37
 Best Local Similarity: 36.77% Mismatches: 113
 Query Match: 27.79% Indels: 77
 DB: 25 Gaps: 16
 US-09-864-291-5 (1-313) x AB267208 (1-1915)

QY 1 MetAlaValaIaSerHisThrGluSerArgGlyValaLeuIleProSerGlyGlu 20
 DB |||||.....
 59 ATGGCGCTCAACAGATCATCTCGAGGCGC---GGCGAGTATCTCATTAACACCGAG 115
 QY 21 SerValLeuIleGlySerGluIleValaLeuGlyPheLeuGlnIleProValaGluSer 40
 DB |||||.....
 116 AGCATCTTAATGCTCTATGATATCATCGGAATCATTAATGACATGAAGAACGTGCCA 175
 QY 41 TyrLeuPheAsnGlyThrIleValaLeuIlePheLeuThrSerTyrValaValaPhe 60
 DB |||||.....
 176 GAAGCTTCAAGAGGACCAAGAAAGGACATGCTACCTTAACCGGATCATCTT 235
 QY 61 ValThrSerHisLeuValaIleAspPheLeuSerPheMetSerProPheGlyLeuMet 80
 DB |||||.....
 236 CTGTCCAAAGGCGC---AAGATGCGCAAGCATGCTCTCATATGATCATTTATATCTCATG 289
 QY 81 SerAspCysThrIleGluIleProIlePheIleProAspThrIleGlyIleGln 100
 DB |||||.....
 290 AAGACTGTGAGTATCAAGACCGCTATTTGTGCAAACTCATCAAGGAAACATGAAAG 349
 QY 101 AlaAlaProGlyGlyIleTyrGluGlyGlnAlaValaPheIleLeuSerPheArgGly 120
 DB |||||.....
 350 GCGGAAGCGGAGTGGCTGGAGAGCTCTCTCTTACAGTTACCTTCAACGAGGAG 409
 QY 121 GlyAlaIleGluPheAlaGluLeuMetValaValaAlaSerAlaAlaArgGly 139
 DB |||||.....
 410 GCGCCATTTGATTCGAGACGCGATGCTGAGTGCATTCACCTTCAAGAGGTGA 469
 QY 140 IleProLeuGlySerValaIleThrPheAspThrSerGlyLeuThrIleIleThrVal 159
 DB |||||.....
 470 GTCCCAAGTGAAGCTATGATGATCTTACATGCGCAAGCGGAGCTTATGATATCCCG 529
 QY 160 ProGlyAlaAlaValaCysSerSerGlnThrProCys---ProAlaTyrProIleVala 178
 DB |||||.....
 530 CCA---GTCCCAATGAATGATACCTTCCCTGCGCTACCTCC 571
 QY 179 TyrGlyProProProGlyTyrThrValGlnProGlyGlyIleThrProProGly 198
 DB |||||.....
 572 TATCCACCGCGCCCACTGATGTC---TATCCAGAGACCGCCCAAG 613
 QY 199 GlyTyrGlyAlaGlnProGlyIleTyrGlyAlaProPro---MetGlyTyr 214
 DB |||||.....
 614 ATGACGAGGCGCATG---GATATGTCGACGCGCCCAACACCGCCTTACCTGAGGCC 667
 QY 215 GlyAlaProProVal---GlyTyrGlyValaPro--- 224
 DB |||||.....
 668 ATGGAACCTCCGCTCAGCGCGCCCGCATGTCCTTCACTCTGACCGCAAGCAAGGCC 727
 QY 225 --- 225
 DB 728 GCAGAAAGACCGCCGACGCTATTAACAACCGCAATCTTCAACCTTCAATGACCC 787
 QY 226 GlyGlyTyrGlyValaProProGlyIleTyrGly--- 236
 DB 788 ACAGACCAAGCGCGCGACCTCTTACTTACCAACCGAAGATGAAGAACCGATGAGCC 847
 QY 237 ---ValProProGlyIle---TyrGlyAlaProPro--- 246
 DB 848 CTCCTGCTCCCTGCTCCGCTCCACCTCATCTCTTACCTTCAACCTTCAACCGGAGCGTGG 907
 QY 247 GlyGlyTyrGlyValaProProGlyIleTyrGlyAlaProProGly---Gly 262
 DB 908 CTGGGCGCTGGGAGGAGGAGGCGCGCTGTGTCCTTCCAGGTCTGATCATTAACAT 967
 QY 263 TyrGlyAlaProProProAlaGlyTyrGlyAlaProProAlaGlyAlaGlnIleLeuProPr 282
 DB |||||.....
 968 TACCAAGAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1024
 QY 282 caaTyrGlyAlaProSerAlaGlyValThrAlaAlaSerHisArgSerMetThr 300
 DB |||||.....
 1025 CAGCTTCCATGCGACCGCGAGCGCAAGTGTGCCAGAGTATCTTCCCTCAAC 1079

RESULT 9

AB870430
 ID AB870430 standard; cDNA, 1885 BP.
 XX
 AC AB870430;
 XX
 DT 27-NOV-2002 (first entry)
 XX
 DE Human bone remodelling gene #87.
 XX
 KM Bone remodelling; osteoporosis; human; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN U66426186-B1.
 XX
 PD 30-JUL-2002.
 XX
 PP 18-JAN-2000; 2000US-0484970.
 XX
 PR 18-JAN-2000; 2000US-0484970.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Jones KA, Volkmutz W, Walker MG,
 XX
 DR WPI: 2002-673014/72.
 XX
 PS A combination of polynucleotides which are co-expressed with genes
 XX known to be involved in bone remodeling and osteoporosis are useful in
 PT an array for the diagnosis of bone remodeling and osteoporosis
 PT associated disorders -
 XX
 PS Claim 1; Column 247-250; 206pp; English.
 XX
 CC The invention relates to a combination comprising a number of
 CC substantially purified and isolated polynucleotides which are
 CC co-expressed with genes known to be involved in bone remodeling and
 CC osteoporosis. The invention is used to diagnose disorders associated
 CC with bone remodeling or osteoporosis. AB870344-AB870512 represent
 CC human bone remodelling genes of the invention.
 XX
 SQ Sequence 1885 BP; 384 A; 610 C; 476 G; 415 T; 0 other;
 Alignment Scores:
 Pred. No.: 3.67e-17 Length: 1885
 Score: 459.50 Matches: 129
 Percent Similarity: 45.43% Conservative: 40
 Best Local Similarity: 34.68% Mismatches: 100
 Query Match: 26.89% Indels: 104
 DB: 24 Gaps: 17
 US-09-864-291-5 (1-313) x AB870430 (1-1885)
 QY 1 MetcLaValaAngInSerHieThGlusErArGylValaLeuIleProSerGlyGlu 20
 DB 61 ATGGCGCTCAACAGAAATCACTCGAAGGC---GGCGAGTGAAGCTCAATTAACACCGAG 117
 QY 21 SerValLeuYsgInGylAspValAspLeuYsPheLeuGlnYsProValGluSer 40
 DB 118 AGCATCTTAATGCTCTTATGATCAGTGAAGCTCACTCAATTCAATGACATGAAGAAAGTCCCA 177
 QY 41 TyrLeuPheAngInGylThyLysGlyThyLeuPheLeuThySerTyArgValaIlePhe 60
 DB 178 GAAGCCTTCAAGGAGCAAGAAAGGCACTGCTTACCTTACCCCTTACCGGTCATCTTT 237
 QY 61 ValThreSerHieLeuValaAspAPrometLeuSerPheMetMetProPheGlyLeuMet 80
 DB 238 CTGTCCAAAGGC-----AAGGATGCCATGACATGCTTCAATGATGCCATTTTATCTCATG 291
 QY 81 SerAspCysThrIleGlnInProIlePheAlaProAsnTyrlleYsgIlyThrIleGln 100
 DB 292 AAAGACTGTGATGATCAAGACCGCCGTTTGTGTGCATTAATCAAGGAAACGATGAG 351

QY 101 AlaAlaProGlyGlyGlyTyrGlnGlnAlaValaPheYsLeuSerPheArgGly 120
 DB 352 GCGGAAGCGGAGAGTGGCTGGAGAGGCTGCTCTTCTTCAAGTGAATTTTCAAGGAGG 411
 QY 121 GlyAlaIleGlnPheAlaGlnLeuMetValaLeuAlaSerAlaAlaIleArgGly 139
 DB 412 GGGCCATTTGATTCGAGACGCGAATGCTCAGATGGCATCTCAAGCTTCAGAGGTGA 471
 QY 140 IleProLeuGlySerValaSerValaSerTyrrPheAspThrSerGlyLeuTyrlleIleThrVal 159
 DB 472 GTCCCAAGTGAACCTAATGCTACTTCAATGACCCAGCGGGGCTAATGCTATGCCCG 531
 QY 160 ProGlyAlaAlaValaCysSerSerGlnThrProCys---ProAlaTyrrProIleValIle 178
 DB 532 CCA-----GTCCCAATGGAATGATACCTTGCCTCTGTGCTTACCC----- 573
 QY 179 TyrGlyProProProPro-----Gly 185
 DB 574 TATCAACCGCCCGCCACCTGATGTTTATCCAGAACCCCGCATGATGAGCGGCGCATGG 633
 QY 186 TyrThrValGlnProGlyGlyTyrrPheProProGlyGlyTyrrGlyAlaGlnProGly 205
 DB 634 TAC---GTGCAAGCCC-----CCACACCGGCTTAC-----CTGGG 666
 QY 206 GlyTyrrGlyAlaPro----- 210
 DB 667 CCATGGAACCTCGGTTGACGCGGCCCATGATGCCCTTCACTGACGCGGAGCAA 726
 QY 210 ----- 210
 DB 727 GGGCGAGAAAGCAGCGCGCCGCTATTACAAACCGAAGCAATCTTCAACAGCTTACAT 786
 QY 211 -----PrometGlyTyrrGlyAlaProProVal 219
 DB 787 GCCCAAGAGCAGCGCGCCGCACTCTTACTTACCCAGGAAATGAAGAACCGAGTA 846
 QY 220 GlyTyrrGlyAlaProProGlyGlyTyrrGlyAlaProProGlyGlyTyrrGlyValaProPro 239
 DB 847 GGC-----CCTCTTGGC-----TCCCTGCTTCCCACTTCTTCACTTAC 891
 QY 240 GlyGlyTyrrGlyAlaProProGlyGlyTyrrGlyAlaProProGlyGlyTyrrGlyAlaPro 259
 DB 892 CCTTCCCATGCGGGGCTGTGTGGGCTTGG-----GAGGGGAGGGGGGCGCT 939
 QY 260 -----ProGlyGlyTyrrGlyAlaProProAla 268
 DB 940 TGTTCCTCCCTCAGATGATCATTAACAATTACCA-----GGAATTACATTTGTG 990
 QY 269 GlyTyrrGlyAlaProProAlaGlyAlaGlnLeuProProAlaTyrrGlyAlaProSer 288
 DB 991 GACCATTA--GGGCGCCCGGCTTGGGAGAG--GTGCGCGCCAGCTTCCATGCCAGGCC 1046
 QY 289 AlaGlyAsnThrAlaAlaSerHieArgSerMetThr 300
 DB 1047 GAGGCCCAAGTGTGCGCCAGGCTTACCTCCATCC 1082
 RESULT 10
 AB84798
 ID AB84798 standard; cDNA; 220895 BP.
 AC AB84798;
 XX
 DT 14-AUG-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #1369.
 XX
 KM Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KM viral infection; parasitic infection; protozoal infection;
 KM fungal infection; sterile inflammatory disease; psoriasis;
 KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KM cardiac reperfusion injury; renal reperfusion injury; AIDS;
 KM adult respiratory distress syndrome; inflammatory bowel disease;
 KM Crohn's disease; ulcerative colitis; periodontal disease;

KM granulocyte activation; chronic inflammation; allergy.
 OS Homo sapiens.
 XX
 PN W0200228999-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 03-OCT-2001; 2001WO-US30821.
 XX
 PR 03-OCT-2000; 2000US-237189P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX
 DR WPI; 2002-435328/46.
 XX
 PT Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -
 XX
 PS Claim 1; SEQ ID NO 1369; 114pp; English.
 XX
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) Gs by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GCA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease, also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published/pct_sequences.
 XX
 SQ Sequence 220895 BP; 53760 A; 54808 C; 55831 G; 56496 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 8 72e-15 Length: 220895
 Score: 451.00 Matches: 84
 Percent Similarity: 63.64% Conservative: 7
 Best Local Similarity: 58.74% Mismatches: 10
 Query Match: 26.39% Indels: 42
 DB: 24 Gaps: 2

QY 174 TyrProIleValIleTyrGlyProProProProGlyTyrThrValGlnProGlyGlyTyr 193
 DB 16999 TTCACGATTAATGTCTATGAGAGCCCACTGACAGA-----TAT 170037
 QY 194 GlyThrProProGlyGlyTyrGlyValGlnProGlyGlyTyrGlyValAProProMetGly 213
 DB 170038 GAGAGCCCACTTCCCGAATACGAGCCCACTGACAGATATGAGCCCACTGATGAG 170097
 QY 214 TyrGlyValAProProValGlyTyrGlyValAProProGlyGlyTyrGlyValAProProGly 233
 DB 170098 AATGAAAGCCCGCTGTGGATACAGAGCTCACTGTGCAATATGAGCCCACTCTT 170157
 QY 234 GlyTyrGlyValAProProGlyGlyTyrGlyValAProProGlyGlyTyrGlyValAProPro 253
 DB 170158 GATACGAGAGCCCACTGACAGATATGAGCCCACTGATGAGAGCCCACT 170217
 QY 254 GlyGlyTyrGlyValAProProGlyGlyTyrGlyValAProProGlyGlyTyrGlyValAPro 273
 DB 170218 CTGAGATATGAGAGCCCACTGATATGAGCCCACTCTGATATGAGAGCCCA 170277
 QY 274 ProAlaGlyValAsnGlyIleValAProProAlaTyrGlyValAProSerAlaGlyValAsnThrAla 293
 DB 170278 CTTGACGAAATGAAAGCCCGCTGCGGATACAGAGCTCACTGATATGAGAGCC 170337
 QY 294 AlaSerHisArgSerMetThrAlaGln-----GlnGlyThrSerLeuProThrThr 310
 DB 170338 AGGCTCAAGAAATCTACAGAGCCAGCTCTGAAAGAGAGGCTTCTTCCCTCTGCC 170397
 QY 311 SerSerSer 313
 DB 170398 TCCTCTTCT 170406
 RESULT 11
 ID ABA67850 standard; DNA; 436 BP.
 XX
 AC ABA67850;
 XX
 XX 01-FEB-2002 (first entry)
 DT
 DE Human foetal liver single exon nucleic acid probe #16155.
 XX
 XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 KW Homo sapiens.
 OS
 XX
 PN W0200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human foetal liver -
 XX
 PS Claim 4; SEQ ID NO 16155; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, CC measuring and displaying gene expression in samples derived from human CC foetal liver. The present sequence is a single exon nucleic acid CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the CC printed specification, but was obtained in electronic format directly CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.

SO Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;

Alignment Scores:

Pred. No.:	1.49e-16	Length:	436
Score:	438.00	Matches:	83
Percent Similarity:	62.86%	Conservative:	5
Best Local Similarity:	59.29%	Mismatches:	42
Query Match:	25.63%	Indels:	10
DB:	22	Gaps:	2

US-09-864-291-5 (1-313) x ABA67850 (1-436)

```

QY 177 ValIleTyrGlyProProProProGlyTyrThrValGlnProGlyGluTyrGlyThrPro 196
DB 3 ATTGCTATGAGCCCACTGCAAGA-----TATGAGCCCA 41
QY 197 ProGluGlyTyrGlyValAlaGlnProGlyGlyTyrGlyValAlaProProMetGlyTyrGlyAla 216
DB 42 CTTCCCGATGAGAGCCCACTGCAAGATGAGCCCACTGATGAAATGAGGC 101
QY 217 ProProValAlaGlyTyrGlyValAlaProProGlyGlyTyrGlyValAlaProProGlyGlyTyrGly 236
DB 102 CCGCGTGGATGATGAGAGCCCACTGAGCATGAGAGCCCACTTGTGATGCGA 161
QY 237 ValProProGlyGlyTyrGlyValAlaProProGlyGlyTyrGlyValAlaProProGlyGlyTyr 256
DB 162 GCCCACTGCAAGATGAGAGCCCACTGATGATGAGAGCCCACTTGTGATAT 221
QY 257 GlyAlaProProGlyGlyTyrGlyValAlaProProAlaGlyTyrGlyValAlaProProAlaGly 276
DB 222 GGAACCCCACTGCGATGATGAGAGCCCACTGCGATGATGAGAGCCCACTGCAAGA 281
QY 277 AsnGluAlaLeuProProAlaTyrGluAlaProSerAlaGlyAsnThrAlaAlaSerHis 296
DB 282 AATGAGAGCCCGCTGCGATGATGAGAGCCCACTGCGATGATGAGAGCCCACTGCAAGA 341
QY 297 ArgSerMetThrAlaGln-----GlnGluThrSerLeuProThrThrSerSerSer 313
DB 342 GAATCTACAGCAGCCCAAGCTCTGAAACGAGGCTTCTCTCCCTGCTGCTCTCTCT 401

```

RESULT 12

AAK42003
ID AAK42003 standard; DNA; 436 BP.

AC AAK42003;

DT 06-NOV-2001 (first entry)

DB Human bone marrow expressed single exon probe SEQ ID NO: 16560.

KM Human; bone marrow expressed exon; gene expression analysis; probe;

XX microarray; cancer; leukemia; lymphoma; myeloma; ss.

OS Homo sapiens.

PN WO200157276-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0633366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI, 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -

XX Example 4; SEQ ID NO: 16560; 658pp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.

SO Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;

Alignment Scores:

Pred. No.:	1.49e-16	Length:	436
Score:	438.00	Matches:	83
Percent Similarity:	62.86%	Conservative:	5
Best Local Similarity:	59.29%	Mismatches:	42
Query Match:	25.63%	Indels:	10
DB:	22	Gaps:	2

US-09-864-291-5 (1-313) x AAK42003 (1-436)

```

QY 177 ValIleTyrGlyProProProProGlyTyrThrValGlnProGlyGluTyrGlyThrPro 196
DB 3 ATTGCTATGAGCCCACTGCAAGA-----TATGAGCCCA 41
QY 197 ProGluGlyTyrGlyValAlaGlnProGlyGlyTyrGlyValAlaProProMetGlyTyrGlyAla 216
DB 42 CTTCCCGATGAGAGCCCACTGCAAGATGAGCCCACTGATGAAATGAGGC 101
QY 217 ProProValAlaGlyTyrGlyValAlaProProGlyGlyTyrGlyValAlaProProGlyGlyTyr 236
DB 102 CCGCGTGGATGATGAGAGCCCACTGAGCATGAGAGCCCACTTGTGATGCGA 161
QY 237 ValProProGlyGlyTyrGlyValAlaProProGlyGlyTyrGlyValAlaProProGlyGlyTyr 256
DB 162 GCCCACTGCAAGATGAGAGCCCACTGATGATGAGAGCCCACTTGTGATAT 221
QY 257 GlyAlaProProGlyGlyTyrGlyValAlaProProAlaGlyTyrGlyValAlaProProAlaGly 276
DB 222 GGAACCCCACTGCGATGATGAGAGCCCACTGCGATGATGAGAGCCCACTGCAAGA 281
QY 277 AsnGluAlaLeuProProAlaTyrGluAlaProSerAlaGlyAsnThrAlaAlaSerHis 296
DB 282 AATGAGAGCCCGCTGCGATGATGAGAGCCCACTGCGATGATGAGAGCCCACTGCAAGA 341
QY 297 ArgSerMetThrAlaGln-----GlnGluThrSerLeuProThrThrSerSerSer 313
DB 342 GAATCTACAGCAGCCCAAGCTCTGAAACGAGGCTTCTCTCCCTGCTGCTCTCTCT 401

```

RESULT 13

AAI48070
ID AAI48070 standard; DNA; 436 BP.

AC AAI48070;

DT 17-OCT-2001 (first entry)

DB Probe #16756 used to measure gene expression in human placenta sample.

KW Probe: microarray; human; placenta; antenatal diagnosis;
 KM Genetic disorder; ss.
 XX Homo sapiens.
 XX OS
 XX MO200157272-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX 30-JAN-2001; 2001WO-US00663.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA
 XX (MOLR-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488897/53.
 DR
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 16756; 654bp; English.
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 CX
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Alignment Scores:
 Pred. No: 1,496-16 Length: 436
 Score: 438.00 Matches: 83
 Percent Similarity: 62.86% Conservative: 5
 Best Local Similarity: 59.29% Mismatches: 42
 Query Match: 25.63% Indels: 10
 DB: 22 Gaps: 2
 US-09-864-291-5 (1-313) x AA148070 (1-436)
 QY 177 ValletTygTygProProProTygTygThrValGlnProGlyGlyTygTygThrPro 196
 DB ::::|||||
 QY 3 ATTGCTATGAGAGCCCACTGACAGGA-----TATGAGAGCCCA 41
 DB ::::|||||
 QY 197 ProGluGlyTygTygAlaGlnProGlyGlyTygTygAlaProPrometGlyTygTygAla 216
 DB ::::|||||
 QY 42 CTTCCCGATACGAGAGCCCACTGACAGATATGAGGCCCAACCGTAGAAGAAAGGC 101
 DB ::::|||||
 QY 217 ProProValGlyTygTygValProProGlyGlyTygTygValProProGlyGlyTygTyg 236
 DB ::::|||||
 QY 102 CCGCTGTGGATACAGAGCTCACTTGGCATATGAGGCCCACTCTTGGATACGA 161
 DB ::::|||||
 QY 237 ValProProGlyGlyTygTygAlaProProGlyGlyTygTygAlaProProGlyGlyTyg 256
 DB ::::|||||
 QY 162 GCCCAGCTGAGATATGAGGCCCACTTGAAGATATGAGGCCCACTCTTGGATAT 221
 DB ::::|||||
 QY 257 GAlaAlaProProGlyGlyTygTygAlaProProAlaGlyTygTygAlaProProAlaGly 276
 DB ::::|||||
 QY 222 GAAAGCCCACTCTCGATATGAGGCCCACTCTCGATATGAGGCCCACTCTGAGAGA 281
 DB ::::|||||
 QY 277 AasnGlnAlaLeuProProAlaTygTygAlaProProAlaGlyTygTygAlaAlaSerHis 296
 DB ::::|||||
 QY 282 AATAAGAGCCCGCTCGCGGATACAGAGCTCACTCTGATACAGAGCCCAAGCCCTCAG 341
 DB ::::|||||
 QY 297 ArgSerMetThrAlaGln-----GlnGlnThrSerLeuProThrThrSerSerSer 313

DB 342 GAATCTACAGACAGCCCAAGCTCTCTGTAAGAAAGAGAGCTTCTTCTCTGCTCTCTCT 401
 RESULT 14
 ID ABS16034 standard; DNA; 436 BP.
 XX ABS16034,
 AC
 XX 19-AUG-2002 (first entry)
 DT
 XX Human genome-derived single exon probe ORF from lung SEQ ID No 16025.
 DE
 XX Human; ds, single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary hemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen's syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease; open reading frame; ORF.
 KW
 XX Homo sapiens.
 OS
 XX MO200186003-A2.
 PN
 XX 15-NOV-2001.
 PD
 XX 30-JAN-2001; 2001WO-US00665.
 PP
 XX 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 PA
 XX (MOLR-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2002-114183/15.
 DR
 XX Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 PT
 XX Claim 4; SEQ ID No 16025; 634bp; English.
 PS
 XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridize at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridization to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one

CC	of 12011 sequences, mentioned in the specification, or encoded by the
CC	probes/open reading frames (ORF). The probes are used for gene
CC	expression analysis, and for identifying exons in a gene, particularly
CC	using human lung derived mRNA and for the study of lung diseases
CC	such as asthma, lung cancer, chronic obstructive pulmonary disease
CC	(COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC	fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC	Niemann-Pick disease, Hermansky-Rudik syndrome, sarcoidosis, pulmonary
CC	haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC	pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic
CC	and hyaline membrane disease. The present sequence is a single exon
CC	probe open reading frame of the invention.
CC	Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences.
XX	
5Q	Sequence 436 BP, 98 A, 146 C, 106 G, 86 T, 0 other;
	Alignment Scores:
	Pred. No.: 1.49e-16 Length: 436
	Score: 438.00 Matches: 83
	Percent Similarity: 62.86% Conservative: 5
	Best Local Similarity: 59.29% Mismatches: 42
	Query Match: 25.63% Indels: 10
	DB: 24 Gaps: 2
US-09-864-291-5 (1-313) x ABS16034 (1-436)	
OY	177 ValIleTYrGIYProProProGIYTYrThValGIaProGIYrGIYThPro 196
Db	3 ATGTCTAATGAGAGCCCACTGAGCA-----TATGAGCCCA 41
OY	197 ProGIaGIYTYrGIYAlaGIaProGIYrGIYTYrGIYAlaProProMeGIYTYrGIYAla 216
Db	42 CTTCCCGATAGAGAGCCCACTGAGATATGAGCCCAACCCGTAAGAAATGAAGC 101
OY	217 ProProValaGIYTYrGIYValProProGIYrGIYTYrGIYValProProGIYrGIY 236
Db	102 CCGCTGTGGATACAGAGCTCACTGTGGATATGAGCCCACTCTGATACGA 161
OY	237 ValProProGIYrGIYTYrGIYAlaProProGIYrGIYTYrGIYValProProGIYrGIY 256
Db	162 GCCCACTGAGAGATATGAGCCCACTGAGATATGAGCCCACTTGATAT 221
OY	257 GIYAlaProProGIYrGIYTYrGIYAlaProProAlaGIYTYrGIYAlaProProAlaGIY 276
Db	222 GGAACCCCACTCTGGATATGAGCCCACTCTGGATATGAGCCCACTGAGGA 281
OY	277 AaGIaAlaLeuProProAlaTYrGIYAlaProSeAlaGIYAnThAlaAlaSerHis 296
Db	282 AATGAAGCCCGCTCGGGATACAGAGCTCACTGTGATACAGAGCCAGCTCAG 341
OY	297 ArgSerMetThrAlaGln-----GInGIYThSerIleuProThrThrSerSer 313
Db	342 GAATCTACAGAGCCAGGCTCTGAAGAAAGAGGCTTCTCTCTCTGCTCTCTTCT 401
RESULT 15	
AL04882	
ID	AL04882 standard; DNA; 7099 BP.
XX	
AC	AA04882;
XX	
DT	21-NOV-2001 (first entry)
XX	
DE	Human reproductive system related antigen DNA SEQ ID NO: 7570.
XX	
KW	Human; reproductive system related antigen; reproductive system disorder;
XX	cancer; gene therapy; de.
XX	
OS	Homo sapiens.
XX	

PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246539.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;

XX
 DR MPI; 2001-465570/50.
 XX
 PT Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 XX
 PS Disclosure; SEQ ID NO 7570; 1297bp + Sequence Listing; English.
 XX
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a genomic sequence encoding a
 CC protein of the invention.
 XX
 SEQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
 Alignment Scores:
 Pred. No.: 5,74e-15 Length: 7099
 Score: 429.50 Matches: 77
 Percent Similarity: 64.34% Conservative: 6
 Best Local Similarity: 59.69% Mismatches: 39
 Query Match: 25.13% Indels: 7
 DB: 22 Gaps: 1
 US-09-864-291-5 (1-313) x AML04882 (1-7099)
 QY 174 TyrProIleValIleTyrGlyProProProProGlyTyrThValGlnProGlyGlyTyr 193
 DB 6719 TTCGAGTTATTGTCTATGAGCCCACTGACAGA-----TAT 6757
 QY 194 GlyThrProProGlyGlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProProMetGly 213
 DB 6758 GAGGCCCACTCCGATAGAGAGCCCACTGAGATATGAGCCCAAGCCCAAGAGA 6817
 QY 214 TyrGlyAlaProProValGlyTyrGlyValProProGlyGlyTyrGlyValProProGly 233
 DB 6818 AATGAGGCCCGCTGTGATACAGACCTCACTGTGATATGAGCCCACTCTT 6877
 QY 234 GlyTyrGlyValProProGlyGlyTyrGlyAlaProProGlyGlyTyrGlyValProPro 253
 DB 6878 GATACGAGGCCCACTGAGATATGAGCCCACTCTGATATGAGCCCACT 6937
 QY 254 GlyGlyTyrGlyAlaProProGlyGlyTyrGlyAlaProProGlyGlyTyrGlyAlaPro 273
 DB 6938 CTGATATGAGAACCACTCTGATATGAGCCCACTCTGATATGAGCCCA 6997
 QY 274 ProAlaGlyAsnGlyAlaLeuProProAlaTyrGlyAlaProSerAlaGlyAsnThrAla 293
 DB 6998 CCTGCAGAAATGAGGCCCGCTGCGGATACAGAGCTCACTGCTGATCAGAGCC 7057
 QY 294 AlaSerHisArgSerMetThrAlaGln 302
 DB 7058 AGGCTCAGGAATCTACAGACCCAG 7084

Search completed: December 16, 2003, 18:07:55
 Job time : 345.166 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - nucleic search, using frame_plus.p2n model

Run on: December 16, 2003, 17:36:35 / Search time 75.9653 Seconds
(without alignments)
1818.631 Million cell updates/sec

Title: US-09-864-291-5
Perfect score: 1709
Sequence: 1 MAVNQSHTSRGALIPSGR.....ASHRSMTAQETSLEPTSSS 313

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DRV=x1p
-Q=/cgn2_1/USPTO.spool_p/US09864291/runat.15122003.160859.23332/app_query.fasta_1.846
-DB=Issued Patents NA -QPMT=faetap -SUPFIX=rn1 -MINMATCH=0.1 -LOOPCT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.ccl
-LIST=45 -DOCALLIGN=200 -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-MODE=LOCAL -OUTPMT=ptco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US09864291@cgn2_1.115@runat.15122003.160859.23332 -NCPG=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELCK=100 -LONGIO
-DEV TIMOUT=120 -MARN TIMOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOB=6 -DELEXT=7

Database: Issued Patents NA.*
1: /cgn2_6/ptodata/2/1na/5A COMB.seq.*
2: /cgn2_6/ptodata/2/1na/5B COMB.seq.*
3: /cgn2_6/ptodata/2/1na/6A COMB.seq.*
4: /cgn2_6/ptodata/2/1na/6B COMB.seq.*
5: /cgn2_6/ptodata/2/1na/6C COMB.seq.*
6: /cgn2_6/ptodata/2/1na/6D COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	459.5	26.9	1885	4	US-09-864-970B-87 Sequence 87, App1
2	227	13.3	2830	2	US-09-010-928B-1 Sequence 1, App1
3	210.5	12.3	2834	2	US-09-010-928B-3 Sequence 3, App1
4	210.5	12.3	4403765	3	US-09-103-840A-2 Sequence 2, App1
5	210.5	12.3	4411529	3	US-09-103-840A-1 Sequence 1, App1
6	206.5	12.1	654	3	US-08-998-416-1144 Sequence 1144, App1
7	205.5	12.0	756	1	US-08-642-255-50 Sequence 50, App1
8	198	11.6	2344	4	US-09-347-878-31 Sequence 31, App1
9	198	11.6	2428	1	US-08-445-050-1 Sequence 1, App1
10	198	11.6	2428	1	US-08-204-691-1 Sequence 1, App1
11	198	11.6	2428	4	US-09-355-295B-2 Sequence 2, App1
12	198	11.6	2487	3	US-08-370-223-12 Sequence 12, App1

13	198	11.6	3018	1	US-08-347-718B-3 Sequence 3, App1
14	198	11.6	3018	1	US-08-482-262-3 Sequence 3, App1
15	198	11.6	3018	6	5200183-1 Patent No. 5200183
16	197.5	11.6	744	4	US-09-183-861-60 Sequence 60, App1
17	197.5	11.6	744	4	US-09-022-765-60 Sequence 60, App1
18	197.5	11.6	744	4	US-09-551-974A-60 Sequence 60, App1
19	195	11.4	810	1	US-08-642-255-60 Sequence 60, App1
20	193	11.3	432	1	US-08-642-255-48 Sequence 48, App1
21	192.5	11.3	2834	2	US-09-010-928B-3 Sequence 3, App1
22	189.5	11.1	11531	1	US-08-068-948A-1 Sequence 1, App1
23	189.5	11.1	11531	1	US-08-442-806-1 Sequence 1, App1
24	189.5	11.1	11531	4	US-09-355-295B-1 Sequence 1, App1
25	178.5	10.4	1995	1	US-08-425-069-3 Sequence 3, App1
26	178.5	10.4	1995	2	US-08-317-844B-3 Sequence 3, App1
27	175.5	10.3	3747	2	US-09-080-897-1 Sequence 1, App1
28	175.5	10.3	3747	3	US-09-323-735-1 Sequence 1, App1
29	175	10.2	4403765	3	US-09-103-840A-2 Sequence 2, App1
30	175	10.2	4411529	3	US-09-103-840A-1 Sequence 1, App1
31	172	10.1	4378	2	US-09-080-897-3 Sequence 3, App1
32	172	10.1	4378	3	US-09-323-735-3 Sequence 3, App1
33	172	10.1	4399	3	US-08-899-595-2 Sequence 2, App1
34	170	9.9	2210	3	US-08-464-700-1 Sequence 1, App1
35	170	9.9	2210	3	US-08-464-700-53 Sequence 53, App1
36	170	9.9	2242	1	US-08-641-627A-37 Sequence 37, App1
37	170	9.9	2427	2	US-08-678-039A-39 Sequence 39, App1
38	168.5	9.9	5822	3	US-08-899-595-4 Sequence 4, App1
39	168.5	9.9	5822	3	US-08-899-595-5 Sequence 5, App1
40	167	9.8	3147	2	US-08-781-802-7 Sequence 7, App1
41	167	9.8	3147	3	US-08-694-078-7 Sequence 7, App1
42	167	9.8	3147	3	US-09-058-260-7 Sequence 7, App1
43	167	9.8	35100	2	US-08-770-379-17 Sequence 17, App1
44	167	9.8	35100	3	US-08-757-669A-17 Sequence 17, App1
45	167	9.8	35100	4	US-09-230-371A-17 Sequence 17, App1

ALIGNMENTS

RESULT 1
US-09-864-970B-87
Sequence 87, Application US/09484970B
Patent No. 6426186
GENERAL INFORMATION:
APPLICANT: Jones, Karen A.
APPLICANT: Volkmuth, Wayne
APPLICANT: Walker, Michael G.
TITLE OF INVENTION: BONE REMODELING GENES
FILE REFERENCE: PB-0014 US
CURRENT APPLICATION NUMBER: US/09/484,970B
CURRENT FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 172
SOFTWARE: PERL Program
SEQ ID NO 87
LENGTH: 1885
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6426186 232968.10CB1
US-09-864-970B-87

Alignment Scores:

Pred. No.: 1.05e-28 Length: 1885
Score: 459.50 Matches: 129
Percent Similarity: 45.43% Conservative: 40
Best Local Similarity: 34.68% Mismatches: 100
Query Match: 26.89% Indels: 104
DB: 4 Gaps: 17

US-09-864-291-5 (1-313) x US-09-864-970B-87 (1-1885)

Qy 1 MetAlaValAsnGlnSerHisrHrguSerArGArgGlyAlaLeuIleProSerGlyGlu 20
Db 61 ATGGCCCTCAACAGATCACTGGAGGCC---GGCGAGTATCTCTCAATACACCGAG 117


```

Db      616 GATATGACCACTGATAGTACCCCAAGCGGCTCAGGTGCAAGGTGCGTTGACCTAGTG 675
Qy      75 MetProPheGlyLeuMetSerAspCysThrIleGlu----- 86
Db      676 CCTCCGGGTCCAGAGTGTAGTGCACCTAGTGTGATCTCGTCTAGTAGTAGTAGACCTA 735
Qy      87 ---GlnProIlePheAlaProAsnTyrIleGlyGlyThrIleGlnAla-----AlaPro 103
Db      736 GTGGAACCTCGTCCC-AGCCCTTAATGTGTGCAAGTGAATCTAGACCTGTGTGTATGCACT 794
Qy      104 GlyGlyGlyTyrTrpGluGlyGlnAlaValPheLeuSerPheArgGlyGlyGlyAlaIle 123
Db      795 GGTGGATCC----- 815
Qy      124 GluPheAlaGlnLeuMetValIleAlaSerAlaAlaIleArgGlyIleProLeuGly 143
Db      816 -----GCTGAGATATCCGCGCGCACTGAGAGCACTGCATCC 851
Qy      144 SerValLeuTyrTrp---PheAspThrSerGlyLeuTyrIleIleThrValProGlyAla 162
Db      852 AGCGGCTCTTAACGACCAAGAGTACAGTGTGAACATAT-----GAACT 896
Qy      163 AlaValCysSerSerGlnThrProCysProAlaTyrProIleValIleTyrGlyProPro 182
Db      897 AGTGAAGAAAGTGA-----CCTTGGACCAAGAGTGTGCTGAGAGCACTAC 944
Qy      183 ProPro-----GlyTyrThrValGlnProGlyGlyTyrGly 194
Db      945 AGCCCAAGTGTAGCTGACCTGTGTGTGACGTGAGCTATGACCAAGAGGTGTAGGA 1004
Qy      195 Thr-----ProProGluGlyTyr-----GlyAlaGlnProGlyGlyTyr--- 207
Db      1005 ACTGTGTAGAGCCGACCAAGAGGTGTACGACCTGTGTGAGCCGACCAAGAGGTGTAGGA 1064
Qy      208 ---GlyAlaProProMetGlyTyr-----GlyAlaProProValGlyTyrGly 222
Db      1065 CCGGTGTGAGCCGACCAAGAGGTGTACGACCTGTGTGAGCTGACCAAGAGGTGTAGGA 1124
Qy      223 ValProProGlyGlyTyrGlyValProProGlyGlyTyr-----GlyValProPro 239
Db      1125 ---CCTGTGTGAGCTGG---CCTGAGGTGTACGACCTGTGTGAGCTGAGACT 1172
Qy      240 GlyGlyTyr-----GlyAlaProProGlyGlyTyr-----GlyValProPro 253
Db      1173 GGAAGTTACGACCTGTGTGAGCTGACCTGAGGTGTACGACCTGTGTGAGACTGAGACT 1232
Qy      254 GlyGlyTyr-----GlyAlaProProGlyGlyTyr-----GlyAlaProPro 267
Db      1233 GGTGTGTACGACCTGTGTGAACTGACCTGAGAGGTGTGACCTGAGAGGCTGAGCA 1292
Qy      268 AlaGlyTyrGlyAlaProProAlaGlyAlaGlnAlaLeuProProAlaTyrGlyAlaPro 287
Db      1293 GGAAGATGTGACCTGTGTGTGTGACCTGTGTGTGAGACTGTGTGTGTGAGACT 1352

```

RESULT 3

US-09-010-928B-3
Sequence 3, Application US/09010928B
Patent No. 5994099

GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V
APPLICANT: Hayashi, Cheryl Y
TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
TITLE OF INVENTION: CODING THEREFOR
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 GATEHOUSE RD, SUITE 500E
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22042

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,928B
FILING DATE: 22-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28977
REFERENCE/DOCKET NUMBER: 1447-109P
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2824 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULAR TYPE: cDNA
HYPOTHEICAL: NO
FEATURE:
NAME/KEY: -
LOCATION: 1..2824
OTHER INFORMATION: /note="Flagelliform DNA sequence
OTHER INFORMATION: taken from 3' region. Stop codon begins at position 2722."
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2724
US-09-010-928B-3

```

Alignment Scores:

```

Pred. No.: 3,63e-08 Length: 2824
Score: 210.50 Matches: 62
Percent Similarity: 51.52% Conservative: 6
Best Local Similarity: 46.97% Mismatches: 33
Query Match: 12.32% Indels: 31
DB: 2 Gaps: 12

```

US-09-864-291-5 (1-313) x US-09-010-928B-3 (1-2824)

```

Qy      180 GlyProProProProGlyTyrThrValGlnProGlyGlyTyrGlyThrProProGluGly 199
Db      1291 GGAAGTTATGACCAAGGA---GGGTGTGAGCTGT-----GGAAGTGAACCTGAGGT 1341
Qy      200 Tyr-----GlyAlaGlnProGlyGlyTyr-----GlyAlaProProMetGly 213
Db      1342 TATGACCTGTGTGAGAGTGTGACCTGAGGTGTATGACCTGTGTGAGCTGAGACTGAG 1401
Qy      214 Tyr-----GlyAlaProProValGlyTyrGlyValProProGlyGlyTyrGlyVal 230
Db      1402 TACGACCTGTGTGTCTGTGTCCAGGTGTATACGA-----CCGGGTGTCTGTGT--- 1452
Qy      231 ProProGlyGlyTyr-----GlyValProProGlyGlyTyr-----GlyAla 244
Db      1453 ---CAGGTGTATCGACCCCGGTGTCTCCGACCAAGAGTATCGACCTGTGGGTCT 1509
Qy      245 ProProGlyGlyTyr-----GlyValProProGlyGlyTyr-----GlyAla 258
Db      1510 GAACTGTGTGTATCGATCTGTGGGTGTGACCTGTGTGTATCGACCTGTGGGTCT 1569
Qy      259 ProProGlyGlyTyr-----GlyAlaProProAlaGlyTyrGlyAlaProProAla 275
Db      1570 GAACCTGTGTATCGGTCTGTGAGGTTCGAGACTGTGTGTATGACCTGTGTGTACT 1629
Qy      276 GlyAlaGlnAlaLeuProProAlaTyrGlyAlaPro 287
Db      1630 GAACCTGTGTATCGACCTGTGTGTGTGAGACT 1665

```

RESULT 4

US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328

GENERAL INFORMATION:
APPLICANT: FLRISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 0.000385 Length: 4403765
Score: 210.50 Matches: 54
Percent Similarity: 56.00% Conservative: 27
Best Local Similarity: 54.00% Mismatches: 2
Query Match: 12.32% Indels: 17
Gaps: 10

US-09-864-291-5 (1-313) X US-09-103-840A-2 (1-4403765)

QY 183 ProProGlyTYrThrValGlnProGlyGlyThrProProGlyGlyTYrGlyAla 202
DB 4122678 CCGCTGGCTACCGCGTTGAACCA-----CCGCCGCGCGCGCGGATAT 4122722
QY 203 GlnProGlyGlyTYrGlyAlaProProMetGlyTYrGlyAlaProProValGlyTYr 222
DB 4122723 GCGCGCGCGGATAT-----CCGCC-----TACCCCGCTACACACCGCGGTACGAC 4122770
QY 223 ValProProGlyGlyTYrGlyValProProGlyGlyTYrGlyValProProGlyTYr 242
DB 4122771 -----CCGCCGCGGTATATGATGCGCGCC-----AGCTATGAC-----CCCCCGCTGGCTAT 4122818
QY 243 GlyAlaProProGlyGlyTYrGlyValProProGlyGlyTYrGlyValProProGly 262
DB 4122819 GGT-----CCACCCCTCGGCTACCCCGCGCACCGCGCGGTGGCGCCACCGCC-----GGC 4122872
QY 263 TyrGlyAlaProProAlaGlyTYrGlyValProProAlaGlyAsnGlyAlaLeuProPro 282
DB 4122873 TATGAC-----CCACCGCTCGGCTATGAC-----CCACCGGTGCGCCCGCGCGGTCAACCA 4122926

RESULT 5
US-09-103-840A-1

Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37RV
US-09-103-840A-1

Alignment Scores:
Pred. No.: 0.000386 Length: 4411529
Score: 210.50 Matches: 54
Percent Similarity: 56.00% Conservative: 27
Best Local Similarity: 54.00% Mismatches: 2
Query Match: 12.32% Indels: 17
Gaps: 10

US-09-864-291-5 (1-313) X US-09-103-840A-1 (1-4411529)

QY 183 ProProGlyTYrThrValGlnProGlyGlyThrProProGlyGlyTYrGlyAla 202
DB 4130483 CCGCTGGCTACCGCGTTGAACCA-----CCGCCGCGCGCGCGGATAT 4130527
QY 203 GlnProGlyGlyTYrGlyAlaProProMetGlyTYrGlyAlaProProValGlyTYr 222
DB 4130528 GCGCGCGCGGATAT-----CCGCC-----TACCCCGCTACACACCGCGGTACGAC 4130575
QY 223 ValProProGlyGlyTYrGlyValProProGlyGlyTYrGlyValProProGlyTYr 242
DB 4130576 -----CCGCCGCGGTATATGATGCGCGCC-----AGCTATGAC-----CCCCCGCTGGCTAT 4130623
QY 243 GlyAlaProProGlyGlyTYrGlyValProProGlyGlyTYrGlyValProProGly 262
DB 4130624 GGT-----CCACCCCTCGGCTACCCCGCGCACCGCGCGGTGGCGCCACCGCC-----GGC 4130677
QY 263 TyrGlyAlaProProAlaGlyTYrGlyValProProAlaGlyAsnGlyAlaLeuProPro 282
DB 4130678 TATGAC-----CCACCGCTCGGCTATGAC-----CCACCGGTGCGCCCGCGCGGTCAACCA 4130721

RESULT 6
US-08-998-416-1144

Sequence 1144, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippaen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Redlschning, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYYP11
TITLE OF INVENTION: AND US8 THEROP
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESSES:
ADDRESSES: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PP/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1144:
SEQUENCE CHARACTERISTICS:

Db 308 CCCGAGACCCGCGGAGACACCCGCTCCAGGCGACCTGGCGGCGCTTCAGAGTCCCGGA 249
Qy 258 AAlaPro-----ProGlyGlyTyrglyAlaPro-----ProAlaGlyTyrglyAlaPro 273
Db 248 GACCCCGGAGACACCCGCTCCAGGCGACCTGGCGGCGCTTCAGAGTCCCGGAACACCC 189
Qy 274 -----ProAlaGlyAlaSerGlyAlaLeuProProAlaTyrglyAlaProSerAla---Gly 290
Db 188 GGTGAGACACCCGCTCCAGGCGCA---CCTGGCGGCGCTTCAGAGTCCCGGAACACCCGCT 132
Qy 291 AenThrAlaAlaSerHisArgSerMetThrAlaGlnGlnGlnThrSerLeuProThrThr 310
Db 131 GG-ACCAGCCGCTCCAGGCGACCTGGCGGCTTCAGAGTCCCGGAACACCCGCTGAGCC 73
Qy 311 SerSerSer 313
Db 72 ACCCGGCTCC 64

RESULT 8

US-09-347-878-31
Sequence 31, Application US/09347878C
Patent No. 6376210
GENERAL INFORMATION:
APPLICANT: Yuan, Chong
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
FILE REFERENCE: 25885-1651
CURRENT APPLICATION NUMBER: US/09/347,878C
CURRENT FILING DATE: 1999-07-06
NUMBER OF SEQ ID NOS: 75
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 31
LENGTH: 2344
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (9)..(2252)
FEATURE:
OTHER INFORMATION: Human cholesterol esterase cDNA
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: M55201/Genbank
US-09-347-878-31

Alignment Scores:

Pred. No.:	3, 01e-07	Length:	2344
Score:	198.00	Matches:	110
Percent Similarity:	39.238	Conservative:	32
Best Local Similarity:	30.394	Mismatches:	127
Query Match:	11.594	Indels:	94
DB:	4	Gaps:	26

US-09-864-291-5 (1-313) x US-09-347-878-31 (1-2344)

Qy 7 HstHrgLusSerArgArgGlyAlaLeuLleProSerGlyGluSerValLeuLysGlnCys 26
Db 1176 TACACTGAGTCC-----TGGGCCGAGACCATCCAGAGAAATGAAGAAAGACTGTG 1229
Qy 27 GluAer-----ValaLeuLysCysPheLeuGlnLysProValGlu----- 39
Db 1230 GTGACCTTGTGAGACCGATGTCTTCTCG--GTGCCACCGAGATGCTCCAGCCAG 1286
Qy 40 -----SerTyLeuPheAlaGlnGlyThrLys 48
Db 1287 CACAGAGCCAAATGCCAAGATGCCAAGACTACCTACTGTTTCCATCCCTCTCGG 1346
Qy 49 GlyThrLeuPheLeuThrSerTyArgValValPheValThrSerHisLeuValaAer 68
Db 1347 ATGCCCGTCTAC-----CCCAATGGGTGGGGCCGACCATGCATGCATGAC 1391
Qy 69 PrometLeuSerPheMetMetProPheGlyLeuMetSer-----AerCys 83
Db 1392 ATTCAGTACGTTTGGGAAAGCCCTTCGCCACCCCGAGGCTACCGGCCCAAGACAG 1451

Qy 84 ThrIleGluInProlIlePheAla-----ProAenTyrlIleGlyThrIleGlnAla 101
Db 1452 ACACTCTTAAGGCAATGATCGCTACTGACACCAATCTTGGCAAA-----ACAGGG 1502
Qy 102 AAlaProGlyGly-----ThpGluGlyAlaValaPheLys 114
Db 1503 GACCCCAACATGGGCACTGGCTGGCCACACACTGGAAA-----CCC 1547
Qy 115 LeuSerPheArgGlyGlyGlyAlaIleGluPheAlaGlnLeuMetValLysAlaAer 134
Db 1548 TACACTACGAAAGAAACAGGGCTACTGAGATCAACAAAGATG-----GGCAGCAGC 1601
Qy 135 AAlaAlaAlaArgGlyIleProLeuGlySerValaenTyrlThrPheAerThr--SerGly 153
Db 1602 TCCATGAAGCGGACCTGAGAACCACTTCCTGGCTACTGAGACCTCACTATCTGCGG 1661
Qy 154 LeuTyrlIleIleThrValProGlyAlaAla--ValCysSerSerGln----- 168
Db 1662 CTGCCCCACATGACCGAACACAGAGGCCACCCCTGTGCCCCCAACAGGAGACTCCGAGGCC 1721
Qy 169 -ThrProCysProAlaTyrlProIleValIleTyrglyProProProGlyTyrlThrVa 188
Db 1722 ACTCCCGTCCCC-----CCAGGGTGACTCCGAGACCGCCCTCGGCCGCCACGGGT 1775
Qy 188 GlnProGlyGlyTyrglyThrProProGlyGly--TyrglyAlaGlnPro----- 204
Db 1776 GACTCCGGGGGGCCCC-CCCGTGGCGGCCACAGGGTGACTCCGGGGCCCCCCCCGGCGCC 1834
Qy 205 -----GlyGlyTyrglyAlaProPromet-----GlyTyrglyAlaProProVa 219
Db 1835 CACGGGTGACTCCGGGGGGCCCCCGTGGCGGCCACAGGGTGACTCCGGGGCCCCCGGT 1894
Qy 219 GlnTyrglyAlaProProGlyGlyTyrglyAlaProProGlyGlyTyrglyAlaProPr 239
Db 1895 G-----CGGCCAAGGTGACTCGGGGGCCCCCCCC-----GTGGCGCC 1933
Qy 239 G--GlyGlyTyrglyAlaProProGlyGlyTyrglyAlaProPro--GlyGlyTyrgly 257
Db 1934 CACGGGTGACTCCGGGGGGCCCCCCCC-----GTGGCGGCCACAGGGTGACTCCGG 1981
Qy 257 ValaProPro-----GlyGlyTyrglyAlaProProAlaGlyTyrglyAlaPr 273
Db 1982 GGGCCCCCGGTGCCCGCCACAGGGTGAACCGGGGGCCCCCCCC-----GTGCC 2029
Qy 273 GProAlaGlyAlaGlnAlaLeuProProAlaTyrglyAlaProSerAlaGlyAenThrAl 293
Db 2030 GCCCAGGGTGACTCCGGGGCCCCCCCC-----GTGGCGGCCACAGGGTGACTCCGG 2080
Qy 293 AAlaSerHisArgSerMetThrAlaGlnGlnThrSer--LeuProThrThrSerSe 312
Db 2081 GGGCCCCCGGTGACCCCGAGGTGACTCCGAGACCGCCCGGTGGCGGCCACAGGGTGA 2140
Qy 312 rSer 313
Db 2141 CTCC 2144

RESULT 9

US-08-445-050-1
Sequence 1, Application US/08445050
Patent No. 5763739
GENERAL INFORMATION:
APPLICANT: Blaseckberg, Lars
APPLICANT: Edlund, Michael
APPLICANT: Hansson, Lennart
APPLICANT: Herrell, Olle
APPLICANT: Lundberg, Lennart
APPLICANT: Stromqvist, Mats
APPLICANT: Toernell, Jan
TITLE OF INVENTION: No. 5763739el Polypeptides
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: White & Case

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/445,050
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/204,691
FILING DATE:
APPLICATION NUMBER: SE 9300686-4
FILING DATE: 01-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9300722-7
FILING DATE: 04-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Steiner Ph.D., Richard J
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-850
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 819-8783
FAX: (212) 354-8113
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2428 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: mammary gland
FEATURE:
NAME/KEY: CDS
LOCATION: 82..2319
OTHER INFORMATION: /product= "bile-salt-stimulated
OTHER INFORMATION: 11pase"
FEATURE:
NAME/KEY: exon
LOCATION: 985..1173
FEATURE:
NAME/KEY: exon
LOCATION: 1174..1377
FEATURE:
NAME/KEY: exon
LOCATION: 1378..1575
FEATURE:
NAME/KEY: exon
LOCATION: 1576..2415
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 151..2316
FEATURE:
NAME/KEY: poly_a_signal
LOCATION: 2397..2402
FEATURE:
NAME/KEY: repeat_region
LOCATION: 1756..2283
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..81
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1756..1788

FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1789..1821
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1822..1854
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1855..1887
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1888..1920
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1921..1953
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1954..1986
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1987..2019
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2020..2052
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2053..2085
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2086..2118
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2119..2151
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2152..2184
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2185..2217
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2218..2250
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 2251..2283
US-08-445-050-1
Alignment Scores:
Pred. No.: 3.14e-07 Length: 2428
Score: 198.00 Matches: 110
Percent Similarity: 39.23% Conservative: 32
Best Local Similarity: 30.39% Mismatches: 127
Query Match: 11.59% Indels: 94
Gaps: 26
US-09-864-291-5 (1-313) x US-08-445-050-1 (1-2428)
Qy 7 HieThrgIuseRgrgAlaLeuIleProSeclYqIuseRValLeuYgInCy 26
Db 1249 TACACCAATCC-----TGGCCCAAGACCTCCAGAGATTAAGAAAGACTGTG 1302
Qy 27 GluAp-----ValApLeuCyPheLeuGInLyProValGlu----- 39
Db 1303 GTGACTTGTGAGACCATGCTCTTCTGTG---GTGCCCAAGATTCCTAGCCAG 1359
Qy 40 -----SerTyleuPheLeuGInLyThrLybLy 48
Db 1360 CACAGACCAATGCCAAGAGTGCACCACTACGCTGTTTCCCTCCCTCCG 1419
Qy 49 GlyThreupheLeuThrseryrArgValValPheValThrsHsleuValAsnAp 68
Db 1420 ATGCCCGTCTAC-----CCCAATGGGTGGGGCCCAACCATGCAGATGAC 1464

QY 69 ProMetLeuSerPheMetCProPheGlyLeuMetSer-----ApcYs 83
Db 1465 ATTCAGTACGTTGGGAGGAGCCCTCCGACCCCGCTACCGCGCCCAAGAG 1524
QY 84 ThrIleGlnGlnProIlePheAla-----ProAenTYrIleLeuGlyThrlIleGlnAla 101
Db 1525 AAGGTCCTAAGGCGCATATGCTCTACGACCACTTGGCAAA-----ACAGGG 1575
QY 102 AlaProGlyGlyGly-----TrpGlyGlyGlnAlaValPheYs 114
Db 1576 GACCCCAACATGGGCGACTCGGCTGCCCCACACACTGGAA-----CCC 1620
QY 115 LeuSerPheArgLyGlyGlyAlaIleGlnPheAlaGlnLeuMetValIleAlaIleSer 134
Db 1621 TACACTACGAAACAGCGGCTACCTCGAGATCACCAAGAAATG-----GGCAGAGC 1674
QY 135 AlaAlaAlaArgGlyIleProLeuGlySerValaIleTYrTrpPheAspThr---SerGly 153
Db 1675 TTCATGAAGCGGAGCTGAGAACTTCTCGCGCTACTGAGACCTCTATCTGGCG 1734
QY 154 LeuTYrIleIleThrValProGlyAlaAla-ValCysSerSerGln-----168
Db 1735 CTGCCCAAGTACCGACCAAGAGGCGACCCCTGTCGCCCCCAAGGGGACTCCAGAGCC 1794
QY 169 -ThrProCysProAlaTYrProIleValIleTYrGlyProProProGlyTYrThyA 188
Db 1795 ACTCCCGTGGCC-----CCAGGGGTGACTCCGAGACCGCCCGTGGCGCCAGGGT 1848
QY 188 IglInProGlyGlyTYrGlyThrProProGlyGly---TYrGlyAlaGlnPro-----204
Db 1849 GACTCGGGGGCCCCC-CCCGTCCCGCCACGGGTGACTCCGGGGCCCCCGTGGCGCC 1907
QY 205 ----GlyGlyTYrGlyAlaProProMet-----GlyTYrGlyAlaProProVa 219
Db 1908 CACGGGTGACTCCGGGGCCCCCGCGTGGCGCCACGGGTGACTCCGGGGCCCCCGCT 1967
QY 219 IglTYrGlyValProProGlyGlyTYrGlyValProProGlyGlyTYrGlyValProPr 239
Db 1968 G-----CGGCCACGGGTGACTCCGGGGCCCCCG-----GTGGCGCC 2006
QY 239 G---GlyGlyTYrGlyAlaProProGlyGlyTYrGlyValProPro---GlyGlyTYrGly 257
Db 2007 CACGGGTGACTCCGGGGCCCCCG-----GTGGCGCCACGGGTGACTCCGG 2054
QY 257 ValaProPro-----GlyGlyTYrGlyAlaProProAlaGlyTYrGlyAlaPr 273
Db 2055 GCGCCCCCGCGTGGCGCCACCGGGTACCGCGGGCCCCCG-----GTGCC 2102
QY 273 oProAlaGlyAenGlyAlaLeuProProAlaTYrGlyAlaProSerAlaGlyAenThrl 293
Db 2103 GCGCACGGGTGACTCCGGGGCCCCCG-----GTGCCCGCCACGGGTGACTCCGG 2153
QY 293 AlaSerHlaArgSerMetThrlaGlnGlnThrSer---LeuProThThrSerSe 312
Db 2154 GCGCCCCCGCGTACCCCAAGGTGACTCCGAGACCGCCCGCGTGGCGCCACCGGTGA 2213
QY 312 rSer 313
Db 2214 CTC 2217

RESULT 10
US-08-204-691-1
Sequence 1, Application US/08204691
Patent No. 5827683
GENERAL INFORMATION:
APPLICANT: Blaesberg, Lars
APPLICANT: Edlund, Michael
APPLICANT: Hansson, Lennart
APPLICANT: Herneil, Olie
APPLICANT: Lundberg, Lennart
APPLICANT: Stromqvist, Mats
APPLICANT: Toernell, Jan
TITLE OF INVENTION: No. 5827683el Polypeptides

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: White & Case
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/204,691
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9300686-4
FILING DATE: 01-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9300722-7
FILING DATE: 04-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Steiner Ph.D., Richard J
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-850
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)819-8783
FAX: (212)354-8113
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2428 base pairs
TYPE: nucleic acid
STRANDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: mammary gland
FEATURES:
NAME/KEY: CDS
LOCATION: 82..2319
OTHER INFORMATION: /product= "bile-salt-stimulated
OTHER INFORMATION: lipase"
FEATURES:
NAME/KEY: exon
LOCATION: 985..1173
FEATURES:
NAME/KEY: exon
LOCATION: 1174..1377
FEATURES:
NAME/KEY: exon
LOCATION: 1378..1575
FEATURES:
NAME/KEY: exon
LOCATION: 1576..2415
FEATURES:
NAME/KEY: mat_peptide
LOCATION: 151..2316
FEATURES:
NAME/KEY: polyA_signal
LOCATION: 2397..2402
FEATURES:
NAME/KEY: repeat_region
LOCATION: 1756..2283
FEATURES:
NAME/KEY: 5'UTR
LOCATION: 1..81
FEATURES:
NAME/KEY: repeat_unit

```

? LOCATION: 1756..1788
? FEATURE: repeat unit
? NAME/KEY: 1789..1821
? LOCATION: 1789..1821
? FEATURE: repeat unit
? NAME/KEY: 1822..1854
? LOCATION: 1822..1854
? FEATURE: repeat unit
? NAME/KEY: 1855..1887
? LOCATION: 1855..1887
? FEATURE: repeat unit
? NAME/KEY: 1888..1920
? LOCATION: 1888..1920
? FEATURE: repeat unit
? NAME/KEY: 1921..1953
? LOCATION: 1921..1953
? FEATURE: repeat unit
? NAME/KEY: 1954..1986
? LOCATION: 1954..1986
? FEATURE: repeat unit
? NAME/KEY: 1987..2019
? LOCATION: 1987..2019
? FEATURE: repeat unit
? NAME/KEY: 2020..2052
? LOCATION: 2020..2052
? FEATURE: repeat unit
? NAME/KEY: 2053..2085
? LOCATION: 2053..2085
? FEATURE: repeat unit
? NAME/KEY: 2086..2118
? LOCATION: 2086..2118
? FEATURE: repeat unit
? NAME/KEY: 2119..2151
? LOCATION: 2119..2151
? FEATURE: repeat unit
? NAME/KEY: 2152..2184
? LOCATION: 2152..2184
? FEATURE: repeat unit
? NAME/KEY: 2185..2217
? LOCATION: 2185..2217
? FEATURE: repeat unit
? NAME/KEY: 2218..2250
? LOCATION: 2218..2250
? FEATURE: repeat unit
? NAME/KEY: 2251..2283
? LOCATION: 2251..2283
?
US-08-204-691-1
Alignment Scores:
Pred. No: 3.14e-07 Length: 2428
Score: 198.00 Matches: 110
Percent Similarity: 39.23% Conservative: 32
Best Local Similarity: 30.39% Mismatches: 127
Query Match: 11.59% Indels: 94
DB: 1 Gaps: 26
US-09-864-291-5 (1-313) x US-08-204-691-1 (1-2428)
Qy 7 HisThrgluserArgrGlyAlaLeuIleProserGlygluserValIleuylGlnCys 26
Db 1249 TACGCCAGATCC-----TGGGCCAGAGACCATCCAGAGAAATTAAGAGAGACTGTG 1302
Qy 27 GluAPr-----ValAPrLeuCyAPheLeuGlnLysProValGlu----- 39
Db 1303 GTGACCTTGAAGACCAATGCTCTCTCTG---GTGCCACCGAGATTGCCCTAGCCGAG 1359
Qy 40 -----SerTyrlleuPheanGlyThrlLysLys 48
Db 1360 CACAGACCAATGCCAAGAGTGCAGACCTACCGCTTACCTGTTTCCATCCCTCTCGG 1419
Qy 49 GJlThrlleuPheLeuThrlSerTyArgrValValPheValThrlSerHlaleuValAsnAPr 68
Db 1420 ATGCCCTCTAC-----CCCAATGGGTGGGGGCGGACCATGCAGATGAC 1464

```

```

Qy 69 ProMetLeuSerPheMetMetProPheGlyLeuMetSer-----APrCys 83
Db 1465 ATTCAGTACCTTTTCGGAAAGCCCTTGGCCACCCCAACGGAGTACCGGCCCAAGACAG 1524
Qy 84 ThrIleGluGlnProIlePheAla-----ProMetTyrlleLysGlyThrlIleGlnAla 101
Db 1525 ACACTCTCTAAGGCCATGATCGCTTACTGACCACTTTCGCANA-----ACAGGG 1575
Qy 102 AlaProGlyGlyGly-----TrpGluGlyGlnAlaValPheLys 114
Db 1576 GACCCCAACATGGGGGAGCTGGGCTGTGCCACACACTGGGAA-----CCC 1620
Qy 115 LeuSerPheArGlyGlyGlyAlaIleGluPheAlaGlnLeuMetValLysAlaAlaSer 134
Db 1621 TACACTACGAAACACAGCGCTACCTGAGATCAACAAAGAAATG-----GCCAGCAGC 1674
Qy 135 AlaAlaAlaAlaProGlyIleProLeuGlySerValAsnTyTrpPheAPrThr-----SerGly 153
Db 1675 TCCATGAAGCGAGACCTGAAGAACCACTTCTGCGCTACTGAACCTTCACTATCTGGCG 1734
Qy 154 LeuTyrlleIleThrlValProGlyAlaAla-ValCysSerSerGln----- 168
Db 1735 CTGCCCAACAGTACCGACAGAGAGCCACCCCTGTGTGCCCAACAGGAGACTCCAGGCC 1794
Qy 169 -ThrlProCyAPrAlaTyTrpProIleValIleTyGlyProProProGlyTyTrlVal 188
Db 1795 ACTCCCTGTGCC-----CCAGGGGTGACTCCGAGACCCGCCCTGTGCCCAACGGGT 1848
Qy 188 GlnProGlyGlyTyTrGlyThrlProProGlyGly-----TyGlyAlaGlnPro----- 204
Db 1849 GACTCCGGGGGCCCCC-CCCGTGCAGCCACAGGGTGACTCCGGGGGCCCCCTGTGCCGC 1907
Qy 205 -GlyGlyTyTrGlyValAPrProMet-----GlyTyTrGlyAlaProProVa 219
Db 1908 CAGGGGTGACTCCGGGGGCCCCCTGTGCCGCCACAGGGTGACTCCGGGGGCCCCCTGT 1967
Qy 219 GJlTyTrGlyValProProGlyGlyTyTrGlyValProProGlyGlyTyTrGlyValProPr 239
Db 1968 G-----CCGCCACAGGGTGACTCCGGGGGCCCCC-----GTGCCGC 2006
Qy 239 o-----GlyGlyTyTrGlyValAPrProProGlyTyTrGlyValProPro-----GlyGlyTyTrGly 257
Db 2007 CAGGGGTGACTCCGGGGGCCCCC-----GTGCCGCCACAGGGTGACTCCGG 2054
Qy 257 ValAPrPro-----GlyGlyTyTrGlyAlaProProAlaGlyTyTrGlyAlaPr 273
Db 2055 CGCCCCCCCCGTGCCGCCACAGGGTGAAGCGGGGCCCCC-----GTGCC 2102
Qy 273 oProAlaGlyAsnGlnAlaLeuProProAlaTyTrGlyAlaProSerAlaGlyAsnThrl 293
Db 2103 GCCCAGCGGGTGACTCCGGGCCCCCCC-----GTGCCGCCACAGGGTGACTCCGG 2153
Qy 293 AlaSerHlaLysSerMetThrlAlaGlnGlnThrlSer-----LeuProThrlThrSerSe 312
Db 2154 GGGCCCCCGGTGAACCCCAAGGGTGACTCCAGACCGCCCGGTGCCGCCACAGGGTGA 2213
Qy 312 rSer 313
Db 2214 CTCC 2217

```

```

RESULT 11
US-09-355-295B-2
/ Sequence 2, Application US/09355295B
/ Patent No. 6525241
GENERAL INFORMATION:
/ APPLICANT: Daijempie, M.
/ APPLICANT: Lundberg, L.
/ APPLICANT: Stromqvist, M.
/ TITLE OF INVENTION: Expression Methods
/ FILE REFERENCE: 1754 SEQUENCE LISTING v2a.txt
/ CURRENT APPLICATION NUMBER: US/09/355,295B
/ CURRENT FILING DATE: 1999-07-18

```

PRIOR APPLICATION NUMBER: PCT/SB99/00648
 PRIOR FILING DATE: 1999-04-21
 NUMBER OF SEQ ID NOS: 5
 SEQ ID NO 2
 LENGTH: 2428
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: CDNA
 FEATURE:
 OTHER INFORMATION: Mammary gland source
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 82..2319
 OTHER INFORMATION: /product= "bile salt-stimulated
 OTHER INFORMATION: 11pase"
 FEATURE:
 NAME/KEY: exon
 LOCATION: 985..1173
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1174..1377
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1378..1575
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1576..2415
 FEATURE:
 NAME/KEY: mat_peptide
 LOCATION: 151..2316
 FEATURE:
 NAME/KEY: polyA_signal
 LOCATION: 2397..2402
 FEATURE:
 NAME/KEY: repeat_region
 LOCATION: 1756..2283
 FEATURE:
 NAME/KEY: 5'UTR
 LOCATION: 1..81
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1756..1768
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1789..1821
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1822..1854
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1855..1887
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1888..1920
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1921..1953
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1954..1986
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1987..2019
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2020..2052
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2053..2085
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2086..2118

FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2119..2151
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2152..2184
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2185..2217
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2218..2250
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2251..2283
 US-09-355-295B-2
 Alignment Scores:
 Pred. No.: 3,14e-07 Length: 2428
 Score: 198.00 Matches: 110
 Percent Similarity: 39.23% Conservative: 32
 Best Local Similarity: 30.39% Mismatches: 127
 Query Match: 11.59% Indels: 94
 DB: 4 Gaps: 26
 US-09-864-291-5 (1-313) x US-09-355-295B-2 (1-2428)
 QY 7 HlHtHrGluSerArgArgGlyAlaLeuIleProSerGlyGluSerValLeuIleGlnCys 26
 DB 1249 TACACCGAGTCC-----TGGGCCGAGACCCATCCCGAGAAATAGAGAAAGACTGTG 1302
 QY 27 GluAap-----ValAspLeuCySPheLeuGlnLysProValGlu----- 39
 DB 1303 GTGGACTTGGAGACCGATGCTCTTCTG--GTGCCACCGAGATTGCCCTACCCAG 1359
 QY 40 -----SerTyrLeuPheAsnGlyThrLysLys 48
 DB 1360 CACAGAGCCAAATGCCAAGATGCCAAGATCCAGCTACCTGTTTCCATCCCTCGG 1419
 QY 49 GlyThrLeuPheLeuThrSerTyrArgValValPheValThrSerHisLeuValAsnAap 68
 DB 1420 ATGCCCGCTTAC-----CCCAATGGGTGGGGCCGACCATGCAGATATAC 1464
 QY 69 PrometLeuSerPheMetSerProPheGlyLeuMetSer-----AapCys 83
 DB 1465 ATTCAGTACGTTTGGGAAGCCCTTCCGACCCCAAGGCTACCGGCCCAAGACAG 1524
 QY 84 ThrIleGlnProIlePheAla-----ProAsnTyrIleLysGlyThrIleGlnAla 101
 DB 1525 ACAGTCTTAAGCCATATGCTGCTACGACCACTTGGCCAA-----ACAAGG 1575
 QY 102 AlaProGlyGlyLys-----TProGlnGlyGlnAlaValPheLys 114
 DB 1576 GACCCCAACATGGGGAGACTCGGCTGTGCCACACACTGGAA-----CCC 1620
 QY 115 LeuSerPheArgLysGlyGlyValIleGlnPheAlaGlnLeuMetValIleAlaIleSer 134
 DB 1621 TACACTACGAAACACGCGGCTACCTGAGATCACCAAGAAATG-----GGCAGACG 1674
 QY 135 AlaAlaAlaArgGlyIleProLeuGlySerValAsnTyrTyrPheAspThr---SerGly 153
 DB 1675 TCCATGAAGCGAGAGCTGAGAACCAATCTCGCGCTACTGACCTCATATCGGG 1734
 QY 154 LeuTyrIleIleThrValProGlyValAlaValCysSerSerGln----- 168
 DB 1735 CTGCCACAGTACGACGACGACGACGACCTTGTGCCCCCAAGGGGACTCCAGAGCC 1794
 QY 169 -ThrProCysProAlaTyrProIleValIleTyrGlyProProProGlyTyrThrVa 188
 DB 1795 ACTCCCGTCCCTCC-----CCAGGGTATCTCCAGACCGCCCTCGTGCAGCCAGGT 1848
 QY 188 IclnProGlyGlnTyrGlyThrProProGlnGly---TyrGlyAlaGlnPro----- 204

Db 1849 GACTCCGGGGCCCCC-CCGTCGCGCCACCGGTGACTCCGGGGCCCCCCTGTCGCC 1907
 GY 205 ----GlyGlyTyrglyAlaProProMet-----GlyTyrglyAlaProProVa 219
 Db 1908 CACGGGTGACTCCGGGGCCCCCCTGTCGCCACCGGGTGACTCCGGGGCCCCCGGT 1967
 GY 219 IGlyTyrglyValProProGlyGlyTyrglyValProProGlyGlyTyrglyValPro 239
 Db 1968 G-----CCGCCACGGGTGACTCCGGGGCCCCC-----GTGCCGCC 2006
 GY 239 o----GlyGlyTyrglyAlaProProGlyGlyTyrglyValProPro----GlyGlyTyrgly 257
 Db 2007 CACGGGTGACTCCGGGGCCCCC-----GTGCCGCCACGGGTGACTCCGG 2054
 GY 257 yAlaProPro-----GlyGlyTyrglyAlaProProAlaGlyTyrglyAlaPr 273
 Db 2055 CCCCCCCCCCGCGCCGCCACCGGTGACCGCGGCCCCCC-----GTGCC 2102
 GY 273 oProAlaGlyAlaGlyAlaLeuProProAlaTyrglyAlaProSerAlaGlyAlaThrAl 293
 Db 2103 GCCCAGGGGTGACTCCGGGGCCCCC-----GTGCCGCCACGGGTGACTCCGG 2153
 GY 293 aAlaSerAlaArgSerMetThrAlaGlnGlnGlnThrSer---LeuProThrThrSerSe 312
 Db 2154 GCGCCCCCCCCCGTACCCCAAGGGTGAATCCGACACCGCCCCCTGTCGCCACCGGTGA 2213
 GY 312 rSer 313
 Db 2214 CTCC 2217

RESULT 12

US-08-370-223-12
 / Sequence 12, Application US/08370223
 / Patent No. 6107026

GENERAL INFORMATION:

APPLICANT: Lange, III, Louis G.

TITLE OF INVENTION: Methods and Reagents for RFLP Analysis

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSER: Allegretti & Witcoff, Ltd.

STREET: 10 South Wacker, Suite 3000

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/370,223

FILING DATE:

CLASSIFICATION: 307

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/053,308

FILING DATE:

APPLICATION NUMBER: 07/730,204

FILING DATE: July 15, 1991

ATTORNEY/AGENT INFORMATION:

NAME: McDonnell, John J.

REGISTRATION NUMBER: 26,949

REFERENCE/DOCKET NUMBER: 91,441

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 2487 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 104..2241
 US-08-370-223-12

Alignment Scores:
 Pred. No.: 3,24e-07 Length: 2487
 Score: 198.00 Matches: 110
 Percent Similarity: 39.23% Conservative: 32
 Best Local Similarity: 30.39% Mismatches: 127
 Query Match: 11.59% Indels: 94
 DB: 3 Gaps: 26

US-09-864-291-5 (1-313) x US-08-370-223-12 (1-2487)

GY 7 HisThrGluSerArgArgGlyAlaLeuIleProSerGlyGluSerValIleuTygGlnCys 26
 Db 1271 TACACCGAGTCC-----TGGGCCACAGACCCATCCAGAGATTAAGAGAGACTGTG 1324
 GY 27 GluAsp-----ValaAspLeuCysPheLeuGlnIlyAsProValGlu----- 39
 Db 1325 GTGACTTGTGAGACCGAGATGCTCTCTCTG---GTGCCACCGAGATTGCCCTAGCCAG 1381
 GY 40 -----SerTyrIleuPheAlaGlyThrIlyLeu 48
 Db 1382 CACAGAGCAATGCCAAGATGCCAAGACCTACCTACCTGTTTCCATCCCTCTCG 1441
 GY 49 GlyThrIleuPheLeuThrSerTyrArgValAlaPheValThrSerHisLeuValAsnAsp 68
 Db 1442 ATGCCCGCTAC-----CCCAATGGGGTGGGGGCCAACCATGCAGATGAC 1486
 GY 69 ProMetLeuSerPheMetMetProPheGlyLeuMetSer-----AspCys 83
 Db 1487 ATTCAGTACGTTTCCGGAAGCCTTCCGCCACCCCAAGGCTACCGGCCCAAGACAG 1546
 GY 84 ThrIleGluIleProIlePheAla-----ProMetTyrIleGlyThrIleGlnAla 101
 Db 1547 ACAGTCTTAAGGCAATGATGCTACTGACCAACTTTCGCAAA-----ACAAGG 1597
 GY 102 AlaProGlyGlyGly-----ThrGluGlyGlnAlaValPheIys 114
 Db 1598 GACCCCAACATGGGAGCTGCGCTGTCGCCACACACTGGAA-----CCC 1642
 GY 115 LeuSerPheArgGlyGlyAlaIleGluPheAlaGlnLeuMetValIlyAlaSer 134
 Db 1643 TACACTACGAAACACGCGCTACTGAGATACCAAGAAATG-----GCCAGCAC 1696
 GY 135 AlaAlaAlaArgGlyIleProLeuGlySerValaMetTyrTrpPheAspThr---SerGly 153
 Db 1697 TCATGAAAGCGAGACCTGAGAACCAACTTCTGCGCTACTGACCTCATTACTTGGCG 1756
 GY 154 LeuTyrIleIleThrValProGlyAlaAla-ValCysSerSerGln----- 168
 Db 1757 CTGCCACAGTACCGACCAAGAGCCACCCCTGTGCCCCCAAGAGGAGACTCCAGGCC 1816
 GY 169 -ThrProCysProAlaTyProIleValIleTyGlyProProProGlyTyTrhVa 188
 Db 1817 ACTCCCGTCCCC-----CACGGGTGACTCCAGACCGCCCCCGTCCGCCACGGGT 1870
 GY 188 IGlnProGlyGlyIuTyrglyThrProProGlyGly---TyrglyAlaGlnPro----- 204
 Db 1871 GACTCCGGGGCCCCC-CCGTCGCGCCACCGGTGACTCCGGGGCCCCCCTGTCGCC 1929
 GY 205 ----GlyGlyTyrglyAlaProProMet-----GlyTyrglyAlaProProVa 219
 Db 1930 CACGGGTGACTCCGGGGCCCCCCTGTCGCCACCGGGTGAATCCGGGGCCCCCGGT 1989
 GY 219 IGlyTyrglyValProProGlyGlyTyrglyValProProGlyGlyTyrglyValProPr 239
 Db 1990 G-----CCGCCACGGGTGACTCCGGGGCCCCC-----GTGCCGCC 2028

QY 239 o---glylytyrglyyalaProProglylytyrglyvalProPro---glylytyrgly 257
 |||||
 DB 2029 CAGGGGTGACTCGGGGGCCCCCCCC-----GTGCGGCCACGGGTGACTCCGG 2076
 QY 257 yalaProPro-----glylytyrglyyalaProProalaglytyrglyalaPr 273
 |||||
 DB 2077 CGCCCCCCCCGTGCGGCCGCCACGGGTGACCGCGGGCCCCCCCC-----GTGCG 2124
 QY 273 oProalaglyanaglyalaLeuProProalaglytyrglyalaProSerAlaglyanThrAl 293
 |||||
 DB 2125 GCCCAGGGGTGACTCGGGGCCGCCCCCC-----GTGCGGCCACGGGTGACTCCGG 2175
 QY 293 alaSerHisArgSerMetThrAlaglnnglnIurThrSer---LeuProThrThrSerSe 312
 |||||
 DB 2176 GGGCCCCCGGTGACCCGCCACGGGTGACTCGAGACCGGCCCGGTGCGGCCGCCACGGGTGA 2235
 QY 312 rSer 313
 |||||
 DB 2236 CTCCT 2239
 RESULT 13
 US-08-347-718B-3
 ; Sequence 3, Application US/08347718B
 ; Patent No. 5696087
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Chu-San
 ; APPLICANT: Tang, Jordan J.N
 ; TITLE OF INVENTION: METHOD FOR REDUCING INTESTINAL ABSORPTION OF
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: Patrea L. Pabst
 ; STREET: 2800 One Atlantic Center
 ; STREET: 1201 West Peachtree Street
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: USA
 ; ZIP: 30309-3450
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/347,718B
 ; FILING DATE: December 1, 1994
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pabst, Patrea L.
 ; REGISTRATION NUMBER: 31,284
 ; REFERENCE/DOCKET NUMBER: OMRF150
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (404) 873-8794
 ; TELEFAX: (404) 873-8795
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3018 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; HYPOTHEetical: NO
 ; ANTI-SENSE: NO
 ; FEATURE:
 ; NAME/KEY: misc. feature
 ; LOCATION: 1..742
 ; OTHER INFORMATION: /Function = "Nucleotides 679 through 2904
 ; OTHER INFORMATION: encode the amino acid sequence for the Human Milk Bile
 ; OTHER INFORMATION: Salt-activated lipase."
 ; US-08-347-718B-3
 Alignment Scores:
 Pred. No.: 4.14e-07 Length: 3018

Score: 198.00 Matches: 110
 Percent Similarity: 39.23% Conserved: 32
 Best Local Similarity: 30.39% Mismatch: 127
 Query Match: 11.59% Indels: 94
 DB: 1 Gaps: 26
 US-09-864-291-5 (1-313) x US-08-347-718B-3 (1-3018)
 QY 7 HsHrGluSerArgArgGlyAlaLeuIleProSerGlyGluSerValIleuGlnCys 26
 |||||
 DB 1837 TACACGAGATC-----TGCGCCAGAGACCATTCAGAGAAATGAAGAGACTGTG 1890
 QY 27 GluAap-----ValAspLeuCybPheLeuGlnIlyProValGlu----- 39
 |||||
 DB 1891 GTGACTTGTGAGACGAGATGCTCTTCTG---GTGCCACCGAGATGTGCCAGCCAG 1947
 QY 40 -----SerTyrIleuPheAnGlyThrIlyAs 48
 |||||
 DB 1948 CACAGAGCCATGCCAGAGATGCCAGAGCTACGCTACTGTCTTCCATCCCTCTCGG 2007
 QY 49 GlyThrLeuPheLeuThrSerTyrArgValPheValIleuSerHisLeuValAsp 68
 |||||
 DB 2008 ATGCCCTCTAC-----CCCAATGGGTGGGGCCGACCATGACATGAC 2052
 QY 69 ProMetLeuSerPheMetMetProPheGlyLeuMetSer-----AapCys 83
 |||||
 DB 2053 ATTCAGTACGTTTGGGAGAGCTTCCGACCCCGAGGGGTACCGGCCCAAGACAGG 2112
 QY 84 ThrIleGlnIleProIlePheAla-----ProAnTyrIleGlyThrIleGlnAla 101
 |||||
 DB 2113 ACAAGTCTTACAGGACATATGCTTACGTGACCACTTGGCAA-----ACAGGG 2163
 QY 102 AlaProGlyGlyIly-----TyrGlnGlyAlaValPheIly 114
 |||||
 DB 2164 GACCCCAATGGGGGACTCGGCTGTGCCACACACTGGAA-----CCC 2208
 QY 115 LeuSerPheArgIlyGlyValAlIleGlnPheAlaGlnLeuMetValIleAlaSer 134
 |||||
 DB 2209 TACATCAGAGAAACAGCGGCTACCTGAGATCACCAAGAGATG-----GGCAGCAGC 2262
 QY 135 AlaAlaAlaArgGlyIleProLeuGlySerValAnTyrTrrPheAspThr---SerIly 153
 |||||
 DB 2263 TCCATGAGCCGAGGCTGAGAACCACTTCTGCGCTCTGACCTGACCTCATCTGAGCG 2322
 QY 154 IeuTyrIleIleThrValProGlyValAlaIleValIleSerSerGln----- 168
 |||||
 DB 2323 GTGCCACAGTACCGACAGAGAGGACCCCTGTGCCCCCAGAGGGAGCTCCAGAGCC 2382
 QY 169 -ThrProCybProAlaTyrProIleValIleTyrIlyProProProGlyTyrThrVa 188
 |||||
 DB 2383 ACTCCCGTGCCTC-----CCAGGGGTGACTCCGAGACCGGCCCGTCCGCGGCT 2436
 QY 188 IglInProGlyIlyTyrIlyThrProProGlnIly---TyrGlyAlaGlnPro----- 204
 |||||
 DB 2437 GACTCGGGGGCCCC-CCGTGCTGCCACCGGTGACTCCGGGGCCCCCGGTGCGGCC 2495
 QY 205 -----GlyIlyTyrGlyAlaProProMet-----GlyTyrGlyAlaProProVa 219
 |||||
 DB 2496 CACGGGTGACTCGGGGGCCCCCGGTGCGGCCACCGGTGACTCCGGGGCCCCCGGT 2555
 QY 219 IglIlyTyrValProProGlyIlyTyrIlyValProProGlyIlyTyrValProPr 239
 |||||
 DB 2556 G-----CCGCCACCGGGTGACTCCGGGGCCCCCCCC-----GTGCGGCC 2594
 QY 239 o---glylytyrglyyalaProProglylytyrglyvalProPro---glylytyrgly 257
 |||||
 DB 2595 CACGGGTGACTCGGGGGCCCCCCCC-----GTGCGGCCACGGGTGACTCCGG 2642
 QY 257 yalaProPro-----glylytyrglyyalaProProalaglytyrglyalaPr 273
 |||||
 DB 2643 CGCCCCCGGTGCGGCCGCCACGGGTGACCGCGGGCCCCCCCC-----GTGCG 2690
 QY 273 oProalaglyanaglyalaLeuProProalaglytyrglyalaProSerAlaglyanThrAl 293

```

Db      2691 GCGCCAGGGTGACTCCGAGGCCCGCCCC-----GNGCACCACAGGGTAGCTCCG 2741
Oy      233 AAlSerHisArgSerMetThrAlaInglngIthrSer---LeuProThrThrSerSe 312
Db      2742 GGCGCCCCCGCGAGCCCGCACAGGGGTGACTCCGAGACCGGCCCGGTGC CGCCACAGGGTGA 2801
Oy      312 rSer 313
Db      2802 CTC C 2805

RESULT 14
US-08-482-262-3
; Sequence 3, Application US/08482262
Patent No. 5821226
GENERAL INFORMATION:
APPLICANT: Wang, Chu-San
APPLICANT: Tang, Jordan J.N
TITLE OF INVENTION: BAL C-TAIL DRUG DELIVERY MOLECULES
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Patrea L. pabet
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Releasee #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,262
FILING DATE: June 7, 1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Pabel, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRP151
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3018 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc. feature
LOCATION: 1..742
OTHER INFORMATION: /Function = "Nucleotides 679 through
OTHER INFORMATION: 2904 encode the amino acid sequence for the Human Milk
US-08-482-262-3

Alignment Scores:
Pred. No.:          4,14e-07      Length:        3018
Score:              198.00       Matches:         110
Percent Similarity: 39.23%     Conservative:    127
Best Local Similarity: 30.38%   Mismatches:     94
Query Match:        11.59%     Indels:         26
DB:                  1           Gaps:            6

US-09-864-291-5 (1-313) x US-08-482-262-3 (1-3018)
Oy      7 HieRrhrgUSeArArgcglvAlaleuileProSeRgIyGlUservAlleuLyggIncYs 26
:::|||||::|||::|

```

Db	1897	TACACCGAGTCC-----TGGGCCCAAGAACCCATCCCAAGAGAAATTAAGAAAGACTGTG	1899
QY	27	GIuApP-----ValAspLeuCyshLeuGlnLysProValGlu-----	39
Db	1891	GTGAGACTTTGAGACCGGATGCTCTTCTGTG---GTGCCCAACGAGATATGCTTAGCCAG	1947
QY	40	-----SerTyrLeuPheAsnGlyThrLysLys	48
Db	1948	CACAGAGCCAAATGCCAAGATGCCAAGACTTAACCTACTGTTTTCCCATCCCTCTCG	2007
QY	49	GlyThrLeuPheLeuThrSerTyrArgValAlaPheValThrSerHLeuValAsnAsp	68
Db	2008	ATGCCCCGTATC-----CCCAATAGGGTGGGGGGCCGACCATGCAATGATATAC	2052
QY	69	ProMetLeuSerPheMetMetProPheGlyLeuMetSer-----AspCys	83
Db	2053	ATTCAAGTACGTTTTTCGGGAAGCCCTTGGCCACCCCAACGGAGCTACCGGCCCAAGACAG	2112
QY	84	ThrIleGluGlnProIlePheAla-----ProAsnTyrIleYsglyThrIleGlnAla	101
Db	2113	ACAGCTCTTAAGGCCATGATGCTGCTTACTGACCAACTTGTCCAA-----ACAAGG	2163
QY	102	AlaProGlyGlyGly-----TrpGluGlyGlnAlaValPheLys	114
Db	2164	GACCCCAACATAGGCGGACTCGGCTGTGCCACACACTGGAA-----CCC	2208
QY	115	LeuSerPheArgLysGlyGlyAlaIleGluPheAlaGlnLeuMetValLysAlaLaser	134
Db	2209	TACACTACGAAAAACAGGCGTACTGAGATATCAACAAAGATG-----GGCAGACG	2262
QY	135	AlaAlaAlaArgGlyIleProLeuGlySerValAsnTyrTrpPheAspThr--SerGly	153
Db	2263	TCCATGAAAGCGAGCCTGAGAACCAACTTCTGTGCTACTGGAACCTCACCATTATCTGGG	2322
QY	154	LeuTyrIleIleThrValProGlyAlaAla-ValCysSerSerGln-----	168
Db	2323	CTGCCCAACATGAGACACACAGAGAGCCACCTGTGCCCCCAACAGGAGACTCGAAGCC	2382
QY	169	-ThrProCyshProAlaTyrProIleValIleTyrGlyProProProGlyTyrThrVal	188
Db	2383	ACTCCCGAGCCCC-----CCAGGGGTGATCTCGAAGACGGCCCCGTGGCCCAAGGT	2436
QY	188	IglnProGlyGlyTyrGlyThrProProGlyGly--TyrGlyAlaGlnPro-----	204
Db	2437	GACTCCGGGGCCCCC-CCCGTGGCCGCCACAGGGTGACTCGGGGGCCCCCCCGTGGCCG	2495
QY	205	-----GlyGlyTyrGlyAlaProPheMet-----GlyTyrGlyAlaProProVal	219
Db	2496	CACGGGTGACTCCGGGGGGCCCCCGTGGCCGCCACAGGGTGACTCGGGGGCCCCCGGT	2555
QY	219	IgLYTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProPr	239
Db	2556	G-----CCGGCCACGGGTGACTCGGGGGCCCCC-----GTGGCGCC	2594
QY	239	G--GlyGlyTyrGlyAlaProProGlyGlyTyrGlyValProPro--GlyGlyTyrGly	257
Db	2595	CACGGGTGACTCGGGGGCCCCC-----GTGGCGCCACGGGTGACTCGG	2642
QY	257	YAlaProPro-----GlyGlyTyrGlyAlaProProAlaGlyTyrGlyAlaPr	273
Db	2643	CGCCCCCCCCCTGCGGCCACAGGGTGAACGGCGGGCCCCCCCC-----GTGCC	2690
QY	273	oProAlaGlyAsnGlnAlaLeuProProAlaTyrGluAlaProSerAlaGlyAsnThrAl	293
Db	2691	GCCACGGGTGACTCGGGGGCCCCC-----GTGGCGCCACAGGGTGACTCGG	2741
QY	293	alaSerHLeaArgSerMetThrAlaGlnGlnIuThrSer---LeuProThrThrSerSe	312
Db	2742	GGCCCCCCCCGTGACCCCAAGGGGTGACTCGAGAACGGCCCCGTGCGCCCAAGGGGA	2801
QY	312	rSer	313
Db	2802	CTCC	2805

RESULT 15
 5200183-1
 / Patent No. 5200183
 / APPLICANT: TANG, JORDAN J.N.; MANG, CHI-SUN
 / TITLE OF INVENTION: RECOMBINANT BILE SALT ACTIVATED LIPASES
 / NUMBER OF SEQUENCES: 22
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/537,426
 / FILING DATE: 12-JUN-1990
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: 504,635
 / FILING DATE: 04-APR-1990
 / APPLICATION NUMBER: 122,410
 / FILING DATE: 19-NOV-1987
 / SEQ ID NO:1:
 / LENGTH: 3018
 / 5200183-1

Alignment Scores:
 Pred. No.: 4.14e-07 Length: 3018
 Score: 198.00 Matches: 110
 Percent Similarity: 39.23% Conservative: 32
 Best Local Similarity: 30.39% Mismatches: 127
 Query Match: 11.59% Indels: 94
 DB: Gaps: 26

US-09-864-291-5 (1-313) x 5200183-1 (1-3018)

QY 7 H1eThqIuSerArgaAglyAlaLeuIleProSerGlyIuSerValIleuYsgInCyS 26
 DB 1837 TACACCGAGTCC-----TGAGCCGAGACCCCACTCCAGAAATAGAAAGAACTGTG 1890
 QY 27 GluAsp-----ValaAspLeuCySPhetLeuGlnIuSerProValGlu----- 39
 DB 1891 GTGACCTTGAAGACCAATGCTCTCTG---GTGCCACCAAGATTGCTTACGCCAG 1947
 QY 40 -----SerTyrlleuPheAsnGlyThrylYsLys 48
 DB 1948 CACAGAGCAATGCCAAGAGTGCACAGACCTACGCTGTTTCCCATCCCTCTCGG 2007
 QY 49 GlYThreupheLeuthSerTyArgValValPheValThrSerHisLeuValAsnAsp 68
 DB 2008 ATGCCCGCTTAC-----CCCAATAGGTGGGGGCCGACCATGCAATGAC 2052
 QY 69 PrometLeuSerPheMetMetProPheGlyLeuMetSer-----AspCyS 83
 DB 2053 ATTCAGTACGTTTTCGGAAAGCCCTTGCCACCCCAAGGGCTACCGGCCCAAGACAGG 2112
 QY 84 ThrIleGluGlnProIlePheAla-----ProAsnTyrlleYsgIYThrIleGlnAla 101
 DB 2113 ACGAGTCTTAAGGCAATGATCGCTACTGACCAACTTGGCAAA-----ACAGGG 2163
 QY 102 AlaProGlyIleGly-----TrpGluGlyGlnAlaValPheLys 114
 DB 2164 GACCCCAACATGGCGGCACTGGCTGTGCCACACACTGGGAA-----CCC 2208
 QY 115 LeuSerPheArgYsgIYsgIYAlaIleGluPheAlaGlnLeuMetValYsAlaIaSer 134
 DB 2209 TACACTACGAAAAACAGCGGCTACCTGAGATCAACAAAGAAATG-----GGCAGCAGC 2262
 QY 135 AlaAlaIaArgGlyIleProLeuGlySerValAsnTyrlTrpPheAspThr-----SerGly 153
 DB 2263 TCCATGAAGGAGGAGCTGAGAACCACTCTGCGCTACTGAGACCTCACTATCTGGCG 2322
 QY 154 LeuTyrlleIleThrValProGlyAlaAla-ValCySerSerGln----- 168
 DB 2323 CTGCCCAAGTGAACCGACAGAGGCCACCCCTGTGTGCCCCCAAGAGGAACTCCGAGGCC 2382
 QY 169 -ThrProCyProIaIaTyrlProIleValIleTyrglyProProProGlyTyThrVa 188
 DB 2383 ACTCCGAGTCC-----CCACGGGTGACTCCGAGACCGCCCGTGGCCGCCACCGGGT 2436

QY 188 IgluProGlyIYgluTyrglyThrProProGluIY----TyrglyAlaGlnPro----- 204
 DB 2437 GACTCCGGGGGCCCC-CCCGTCCCGCCACAGGGTGACTCCGGGGGCCCCCGATGCCGCC 2495
 QY 205 ----GlyIYTyrglyAlaProProMet-----GlyTyrglyAlaProProVa 219
 DB 2496 CACGGGTGACTCCGGGGGCCCCCGGTGCGGCCACAGGGTGACTCCGGGGGCCCCCGGT 2555
 QY 219 IglYTyrglyValProProGlyIYTyrglyValProProGlyIYTyrglyValProPr 239
 DB 2556 G-----CGGCCACGGGTGACTCCGGGGGCCCCCG-----GTGCCGCC 2594
 QY 239 G---GlyIYTyrglyAlaProProGlyIYTyrglyValProPro---GlyIYTyrgl 257
 DB 2595 CACGGGTGACTCCGGGGGCCCCCG-----GTGCCGCCACAGGGTGACTCCGG 2642
 QY 257 ValaProPro-----GlyIYTyrglyAlaProProAlaGlyTyrglyAlaPr 273
 DB 2643 CGCCCCCGCGTGCAGCGGGTGAACCGCGGGCCCCCG-----GTGCC 2690
 QY 273 OProAlaGlyAsnGluAlaLeuProProAlaTyrglyAlaProSerAlaGlyAsnThra 293
 DB 2691 GCCCAGGGGTGACTCCGGGGGCCCCCG-----GTGCCGCCACAGGGTGACTCCGG 2741
 QY 293 AlaSerHisArgSerMetThraGlnGlnIuThrSer---LeuProThrThrSerSe 312
 DB 2742 GGCCTCCCGCGTGAACCCCAAGGGTGACTCCGAGACCGCCCGCGTGCCTCCACAGGGTGA 2801
 QY 312 rSer 313
 DB 2802 CTCC 2805

Search completed: December 16, 2003, 23:51:58
 Job time : 11:55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - nucleic search, using frame_p1n model

Run on: December 16, 2003, 17:58:55 / Search time 306.146 Seconds
(without alignments)
3398.008 Million cell updates/sec

Title: US-09-864-291-5
Perfect score: 1709
Sequence: 1 MAVNQHSHSRKGLIPSGR.....ASHRSMTAQOETSLPTSSS 313

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2201672 seqs, 1661799599 residues
Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-Q=/cg2_1/USPTO.epool.p/US098644291/runat.15122003.160900.23360/app.query.fasta.1.846
-DB=Published Applications NA -QFMT=fastap -SUFFIX=emb -MINMATCH=0.1
-LOOPCT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62
-TRANS=human40.cdt -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODB=LOCAL -OUTFMT=pct -FORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US098644291@cgn.1.1.108@runat.15122003.160900.23360
-NCPU=6 -ICPU=3 -NO MMAP -LARGESUBSTRY -NEG SCORES=0 -WAIT -DSBLOCLCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database: Published Applications NA:
1: /cg2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cg2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cg2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cg2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cg2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cg2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cg2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cg2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cg2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cg2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cg2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cg2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cg2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cg2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cg2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cg2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cg2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cg2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES		
Result No.	Query Match Length DB ID	Description

1	543	31.8	467	11	US-09-764-891-1230	Sequence 1230, Ap
2	438	25.6	436	9	US-09-864-761-31561	Sequence 31561, Ap
3	429.5	25.1	7099	11	US-09-764-891-7570	Sequence 7570, Ap
4	353.5	20.7	432	11	US-09-918-995-33536	Sequence 33536, A
5	347	20.3	321	11	US-09-764-891-1356	Sequence 1356, Ap
6	320.5	18.8	409	10	US-09-960-352-6984	Sequence 6984, Ap
7	241	14.1	593	13	US-10-027-633-290832	Sequence 290832,
8	241	14.1	593	14	US-10-027-633-290832	Sequence 290832,
9	239	14.0	511	13	US-10-029-386-24981	Sequence 24981, A
10	218	12.8	1743	10	US-09-887-576-788	Sequence 788, App
11	216	12.6	2016	10	US-09-938-842A-2004	Sequence 2004, Ap
12	215	12.5	1812	10	US-09-925-300-724	Sequence 724, App
13	214	12.5	2176	13	US-10-007-926A-215	Sequence 215, App
14	211	12.3	439	11	US-09-918-995-4554	Sequence 4554, Ap
15	209.5	12.3	2709	15	US-10-156-761-3543	Sequence 3543, Ap
16	209.5	12.3	9025608	15	US-10-156-761-1	Sequence 1, Appl
17	206	12.1	511	13	US-10-029-386-24981	Sequence 24981, A
18	206	12.1	1674	15	US-10-156-761-3052	Sequence 3052, Ap
19	206	12.1	9025608	15	US-10-156-761-1	Sequence 1, Appl
20	205.5	12.0	1909	13	US-10-094-748-1166	Sequence 1166, Ap
21	205.5	12.0	2709	15	US-10-156-761-3543	Sequence 3543, Ap
22	204	11.9	541	13	US-10-029-386-11265	Sequence 11265, A
23	203	11.9	3760	10	US-09-887-576-458	Sequence 458, App
24	202	11.8	534	10	US-09-938-842A-968	Sequence 968, App
25	202	11.8	731	9	US-09-770-149-103	Sequence 103, App
26	198	11.6	2428	10	US-09-969-347-220	Sequence 220, App
27	198	11.6	2428	11	US-09-418-176-1	Sequence 1, Appl
28	198	11.6	3004	9	US-09-873-367C-446	Sequence 446, App
29	197.5	11.6	744	9	US-09-874-923-60	Sequence 60, Appl
30	197.5	11.6	3138	15	US-10-156-761-2538	Sequence 2538, Ap
31	195.5	11.4	41	15	US-10-156-761-2538	Sequence 3570, Ap
32	193	11.3	1869	15	US-10-156-761-2538	Sequence 1092, Ap
33	191	11.2	540	10	US-09-938-842A-1092	Sequence 647, App
34	190	11.1	80557	16	US-10-080-170-647	Sequence 2531, Ap
35	184	10.7	2430	15	US-10-156-761-7526	Sequence 7526, Ap
36	182.5	10.7	1733	15	US-10-128-714-7526	Sequence 7527, Ap
37	182	10.6	291	11	US-09-764-891-7521	Sequence 263, App
38	182	10.6	507	11	US-09-770-961-253	Sequence 2170, Ap
39	178.5	10.4	1958	10	US-09-880-107-2170	Sequence 207, App
40	178.5	10.4	1958	13	US-10-007-926A-107	Sequence 118, App
41	178.5	10.4	1958	13	US-10-210-120-118	Sequence 225, App
42	178.5	10.4	2108	10	US-09-962-833-225	Sequence 19944, A
43	178.5	10.4	2182	13	US-09-814-351-19944	Sequence 52, Appl
44	178.5	10.4	2468	13	US-10-210-120-52	Sequence 9949, Ap
45	178.5	10.4	8664	15	US-10-198-846-9949	

ALIGNMENTS

RESULT 1
US-09-764-891-1230
Sequence 1230, Application US/09764891
Publication No. US2003007808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764, 891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1230
LENGTH: 467
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (421)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1230
Alignment Scores:

Pred. No.: 2,286-46 Length: 467
 Score: 543.00 Matches: 104
 Percent Similarity: 59.46% Conservative: 12
 Best Local Similarity: 62.28% Mismatches: 38
 Query Match: 31.77% Indels: 13
 DB: 11 Gaps: 4

US-09-864-291-5 (1-313) x US-09-764-891-1230 (1-467)

```

QY 118 ArglySGlyGlyAlaIleGlnPheAlaGlnMetVallyValAlaSerAlaAla 137
DB 3 AGAAATGAGAGGCAATGAAATTCACCACTGATGAGGAAAGCTGCTGCTGCTGCC 62
QY 138 ArgGlyIleProLeuGlySerValIleNTyrTrpPheAspThrSerGlyLeuTyrllelle 157
DB 63 CGAGATTCCTCCACTTGAACCTTAATGACTGTTCACTCTATGAGATTTATTAATT 122
QY 158 ThrValProGlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrProIleVal 177
DB 123 ACTGGGGAAGG---AATATGCACTCCACAGATGCTGT-----TCAGTTAAT 170
QY 178 IleTyGlyProProProProGlyTyrThrValGlnProGlyIleTyGlyThrProPro 197
DB 171 GCTATGAGAGCCCACTGACAGAA-----TATGAGCCCACT 209
QY 198 GlnGlyTyrGlyAlaGlnProGlyTyrGlyAlaProProMetGlyTyrGlyAlaPro 217
DB 210 CCGGATAGAGAGCCCACTGACAGATTTGAGGCCCACTGAGAAATGAGAGCCCG 269
QY 218 ProValGlyTyrGlyValaProProGlyTyrGlyValaProProGlyTyrGlyVala 237
DB 270 CCTGTGGGATACAGAGCCCTCACCTGTGAGATTTGAGAGCCCACTTGTGATACGAGAGCC 329
QY 238 ProProGlyGlyTyrGlyAlaProProGlyTyrGlyValaProProGlyTyrGly 257
DB 330 CCACTGAGAGATGAGAGCCCACTCTAGATTTGAGAGCCCACTTTTGTGATATGAG 389
QY 257 ValaProProGlyTyrGlyValaProProAlaGlyTyrGlyAlaProProAlaGly 277
DB 390 AACCCCACTCTCGAGATTTGAGAGCCCACTTTCGATTTGAGAGCCCACTGACGAGAAA 449
QY 277 ngIuaIaleuProProAla 283
DB 450 TGAAGGCC---CCGCTTGC 465
  
```

RESULT 2
 US-09-864-761-31561
 / Sequence 31561, Application US/09864761
 / Patent No. US20020048763A1
 / GENERAL INFORMATION:
 / APPLICANT: Penn, Sharron G.
 / APPLICANT: Rank, David R.
 / APPLICANT: Hanzel, David K.
 / APPLICANT: Chen, Wensheng
 / TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 / FILE REFERENCE: Aecm1ca-X-1
 / CURRENT APPLICATION NUMBER: US/09/864,761
 / PRIOR FILING DATE: 2001-05-23
 / PRIOR APPLICATION NUMBER: US 60/180,312
 / PRIOR FILING DATE: 2000-02-04
 / PRIOR APPLICATION NUMBER: US 60/207,456
 / PRIOR FILING DATE: 2000-05-26
 / PRIOR APPLICATION NUMBER: US 09/632,366
 / PRIOR FILING DATE: 2000-08-03
 / PRIOR APPLICATION NUMBER: GB 24263.6
 / PRIOR FILING DATE: 2000-10-04
 / PRIOR APPLICATION NUMBER: US 60/236,359
 / PRIOR FILING DATE: 2000-09-27
 / PRIOR APPLICATION NUMBER: PCT/US01/00666
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00667
 / PRIOR FILING DATE: 2001-01-30

/ PRIOR APPLICATION NUMBER: PCT/US01/00664
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00669
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00665
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00668
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00663
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00662
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00661
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00670
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: US 60/236,667
 / PRIOR FILING DATE: 2000-09-21
 / PRIOR APPLICATION NUMBER: US 09/608,408
 / PRIOR FILING DATE: 2000-06-30
 / PRIOR APPLICATION NUMBER: US 09/774,203
 / PRIOR FILING DATE: 2001-01-29
 / NUMBER OF SEQ ID NOS: 49117
 / SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
 / SEQ ID NO 31561
 / LENGTH: 436
 / TYPE: DNA
 / ORGANISM: Homo sapiens
 / FEATURE:
 / OTHER INFORMATION: MAP TO 299716.4
 / OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
 / OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
 / OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
 / OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
 / OTHER INFORMATION: EST HUMAN HIT: AM169960.1, EVALUATE 8.00e-11
 / OTHER INFORMATION: SWISSPROT HIT: P18616, EVALUATE 3.00e-13
 / US-09-864-761-31561

Alignment Scores:
 Pred. No.: 1,026-35 Length: 436
 Score: 438.00 Matches: 83
 Percent Similarity: 62.86% Conservative: 5
 Best Local Similarity: 59.29% Mismatches: 42
 Query Match: 25.63% Indels: 10
 DB: 9 Gaps: 2

US-09-864-291-5 (1-313) x US-09-864-761-31561 (1-436)

```

QY 177 ValIleTyGlyProProProProGlyTyrThrValGlnProGlyIleTyGlyThrPro 196
DB 3 ATTGCTATATGAGAGCCCACTGACAGAA-----TATGAGCCCA 41
QY 197 ProGlnGlyTyrGlyAlaGlnProGlyTyrGlyAlaProProMetGlyTyrGlyAla 216
DB 42 CTTCCGATACGAGAGCCCACTGACAGATTTGAGAGCCCACTGAGAAATGAGAGCC 101
QY 217 ProProValGlyTyrGlyValaProProGlyTyrGlyValaProProGlyTyrGly 236
DB 102 CCGCTGTGGATACAGAGCCCTCACCTGTGAGATTTGAGAGCCCACTTGTGATACGAG 161
QY 237 ValaProProGlyTyrGlyValaProProGlyTyrGlyValaProProGlyTyr 256
DB 162 GCCCACTGAGAGATTTGAGAGCCCACTCTAGATTTGAGAGCCCACTTGTGATAT 221
QY 257 GlnAlaProProGlyTyrGlyValaProProAlaGlyTyrGlyAlaProProAlaGly 276
DB 222 GAAACCCCACTCTCGAGATTTGAGAGCCCACTCTCGATTTGAGAGCCCACTGACAG 281
QY 277 AengIuaIaleuProProAlaTyrGlyAlaProSerAlaGlyAanThrAlaAlaSerHis 296
DB 282 AATGAGAGCCCGCTTGGGATACAGAGCCCTCACCTGTGATACAGAGAGCCCTCAG 341
QY 297 ArgSerMetThrAlaGln-----GlnGlnThrSerLeuProThrThrSerSer 313
  
```

```

Db      342 GATCTACGACGAGCCCAAGGCTCTCTGAAGAAGAGGCTTCTTTCCCTCCTGCTCTCTCT 401
                               |||||
RESULT 3
US-09-764-891-7570
; Sequence 7570, Application US/09764891
; Publication No. US2003007780BA1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7570
LENGTH: 7099
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-891-7570

Alignment Scores:
Pred. No.:          2,4e-33           Length:          7099
Score:              429.50           Matches:         77
Percent Similarity: 64.348            Conservative:    6
Best Local Similarity: 59.69%        Mismatches:     39
Query Match:        25.13%           Indels:         7
DB:                 11               Gaps:           1

US-09-864-291-5 (1-313) X US-09-764-891-7570 (1-7099)

QY      174 TYRProIleValIleTYrGLyProPProPProGLyTYrThrValGIaInProGLyGIuTYr 193
Db      6719 TTCCAGATTATGTCTATGAGAGGCCCACTGGAGGA-----TAT 675
QY      194 GIlyThrProPProGLyGIlyTYrGLyAlaGIaInProGLyGIlyTYrGLyAlaProBromecGIy 213
Db      6758 GGAGGCCCACTCCCGGATACGAGAGGCCCACTGGAGATATGAGACCACCACTTAGGA 681
QY      214 TYrGLyAlaProProValIGLyTYrGLyValIProPProGLyGIlyTYrGLyValIProProGLy 233
Db      6818 AATGAAGCCCCCTGTGGATACAGAGCTCACCTGTGCAGATATGAGAGCCCACCTCTT 687
QY      234 GIlyTYrGLyValIProPProGLyGIlyTYrGLyAlaProPProGLyGIlyTYrGLyValIProPro 253
Db      6878 GGATACGAGAGCCCACTGGAGATATGAGAGGCCCACTCTTGATATGAGAGCCCACCT 693
QY      254 GIlyGIlyTYrGLyAlaProPProGLyGIlyTYrGLyAlaProProAlaGIlyTYrGLyAlaPro 273
Db      6938 CTTCGATATGAAACCCCACTCTCGGATATGAGAGCCCACTCTCGATATGAGAGCCCA 699
QY      274 ProAlaGIyaenGIuaIaleuProProAlaTYrGLuAlaProSeAlaGIyaenThzAla 293
Db      6998 CCGGAGGAAATGAAGAGCCCGCTGGGGATACAGAGCTCACCTGGCATCAGAGAGC 705
QY      294 AlaSerHisArgSerMetThrAlaGIu 302
Db      7058 AGCCTCAGGATCTTACGACGAGCCAG 7084

RESULT 4
US-09-918-995-33536
; Sequence 33536, Application US/09918995
; Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID LIBRARIES OBTAINED
FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20

```

```

NUMBER OF SEQ ID NOS: 38054
SOFTWARE: PatSeq for Windows Version 3.0
SEQ ID NO 33536
LENGTH: 432
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1) .. (432)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-33536

Alignment Scores:
Pred. No.: 4.01e-27 Length: 432
Score: 353.50 Matches: 71
Percent Similarity: 68.35% Conservative: 24
Best Local Similarity: 51.08% Mismatches: 41
Query Match: 20.68% Indels: 3
DB: 11 Gaps: 2

US-09-864-291-5 (1-313) x US-09-918-995-33536 (1-432)
QY 1 MetalValAlaEngInSerHisThrGluSerArgArgValAlaLeuIleProSerGlyGlu 20
DB 18 ATGGCGCTCTCAAGAAAGATCTCTGGAGGCGC---GGCGAGTGAATGTCAATACCGGAG 74
QY 21 SerValLeuYsgInGlyGluSerValAlaPheLeuCySPheLeuGlnIleProValGlySer 40
DB 75 AGCATCTTAATGGCGTATATCATCGTGAACCTCACTTCAATGAATGAAGACGTGCA 133
QY 41 TyrLeuPheLeuGlyThrIleYsgIlyThrIlePheLeuThrSerTyrArgValAlaPhe 60
DB 135 GAAGCTTCAAGAGGACCAAGAAAGGCACTGCTCACTTACCCCTTACCGGGTATCTTT 195
QY 61 ValIleSerHisIleValLeuAlaPheProMetLeuSerPheMetLeuProPheGlyLeuMet 80
DB 195 CTGGCCAAAGGCGC-----AAGATGCCATGACGTCTCTCAATGATGCACTTATCTCAAG 248
QY 81 SerAspCySPheThrIleGluGlnProIlePheAlaProAsnTyrIleYsgIlyThrIleGln 100
DB 249 AAGAGCTGTAGATCAAGCAGCCCGTATTTGGTGCMAACTACATCAAGGAAACGTGAAG 308
QY 101 AlaAlaProGlyGlyGlyTyrGluGlyGlnAlaValPheIleLeuSerPheArgIleGly 120
DB 309 GCGGAACGGGAGGTGGCTGGGAGGCTGCTGCTCTCAAGTGAACCTTCAACGGCAGGG 366
QY 121 GlyAlaIleGluIlePheAlaGlnLeuMetValIleValAlaIleSerAlaAlaAlaArgGly 139
DB 369 GGGCCCATGATGAGTTCCGAGACAGCGGAGTCCAGCGGAGCATCTCAAGCCTCCAGAGGT 425

RESULT 5
US-09-764-891-1356
Sequence 1356, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1356
LENGTH: 321
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (316)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (321)

```

! OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1356

US-09-764-891-1356

Alignment Scores:

Pred. No.:	1,27e-26	length:	321
Score:	347.00	Matches:	69
Percent Similarity:	81.90%	Conservative:	17
Best Local Similarity:	65.71%	Mismatches:	18
Query Match:	20.30%	Indels:	1
DB:	11	Gaps:	0

US-09-864-291-5 (1-313) X US-09-764-891-1356 (1-321)

Qy	11	ArgArgGlyValAlaLeuLeuLeuProSerGlyGlyLeuSerValLeuValGlnCysGlyAlaAspValAsp	30
Dd	3	CGCGCGGAGACCTCAATCCCTTAACGGTAGAAAGTCTTGAAGCGGTCTCCGAATATGAG	62
Qy	31	LeuCyAPheLeuGlnIuProValIguSerTyrLeuPheAsnGlyThrIleValGlyThr	50
Dd	63	CTCTCCCTTCCCAACGAGATACGAAGGCTCAATGCTTTAGTGATGAGAAACAGAAACA	122
Qy	51	LeuPheLeuThrSerTyrArgValValPheValThrSerHisLeuValAsnAspProMet	70
Dd	123	TTGTTTTCGACTTCATACCGGGATGATTTTCATATCAATCAATGCTCATACAGATCCATG	182
Qy	71	LeuSerPheMetMetProPheGlyLeuMetSerAspCysThrIleGlnGlnProIlePhe	90
Dd	183	TTGCTCTTTATGATGACATTTGATCTGATGACGAACCTCACTGTGAACCAACGATATTT	242
Qy	91	AlaProAsnTyrIleIleValGlyThrIleGlnAlaIleProGlyValGlyTyrGlu-GlyGly	110
Dd	243	GCTGCAAACTTCATTAAAGGAACTATTCAGCGAGCTTCATATGAGGAGCTGGGAAAGACA	302
Qy	110	AlaValAlaPheIle	114
Dd	303	AGCTACTTTTAAA	315

RESULT 6

```

US-09-960-352-6984
; Sequence 6984, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagesan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 6984
; LENGTH: 409
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 30-LIB3058-024-01-K1-H9
US-09-960-352-6984

```

Alignment Scores:	
Pred. No.:	8.56e-24
Score:	320.50
Percent Similarity:	67.18%
Best Local Similarity:	48.05%
Query Match:	18.75%
DB:	10
Length:	40
Matches:	64
Conservative:	24
Mismatches:	40
Indels:	1
Gaps:	2

US-09-864-291-5 (1-313) X US-09-960-352-6984 (1-409

```

Oy      1 MetAlaValAsnGlnSerHisThrGluSerArgAGGlyAlaLeuIleProSerGlyGlu 20
          |||||:::|||||:::|||||:::
Db      23 ATGGCTTCATCAAGACCACTCGAGGGC---GGCGAGTGCATCTGTCAACACACCGAG 75

```

QY 21 SerValLeuMetGlnGlySerGlnAspValAlaSerPheCysPheLeuGlnIysProValGlnSer 40
 Db 80 AGCATCTGATGTCCTATGACCATGTAGAGCTTACGTTCACTGACATGACATGAGAGATATGCA 139
 QY 41 TyrLeuPheAlaGlyThrIleTyrGlyIhrLeuPheLeuThrSerTyrArgValValPhe 60
 Db 140 GAGGCTTCAAGGGACCAAGAAAGGACACCGTACCTTACCCGATCCGGATCATTT 199
 QY 61 ValThrSerThrIleValAlaAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 Db 200 CTCTCCAAAG-----GCGAGGGATGCTCATGACGTCCTTCATGATGCCCTTCTATCTATG 253
 QY 81 SerAspCysThrIleGlnIleProIlePheAlaProAsnTyrIleTyrGlyThrIleGln 100
 Db 254 AAGACCTGTGAGATTCAGACAGCGCTGTGTGTGGGCGAACTACATCACTCAAGAGGACGTAA 313
 QY 101 AlaIleProGlyValGlyTyrPheGlnIleGlnAlaValPheIleLeuSerPheArgValGly 120
 Db 314 GCCGAAGACGAGAGGTGTGCGAGAAAGCTCTGACTACTCAAGTTGACCTTTATGTTGGGG 373
 QY 121 GlyAlaIleGluPheAlaLeuMetValIys 131
 Db 374 GGCGCCATCATATGACACGCGATGTACAG 406

RESULT 7
ms-10-07

```

US-10-027-632-290832
/ Sequence 290832, Application US/10027632
/ Publication No. US20030204075A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ POLYMORPHISMS IN THE HUMAN GENOME
/ TITLE OF INVENTION: Polymorphisms in the Human Genome
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 290832
/ LENGTH: 593
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-290832

```

Pred. No.:	1.6e-15	length:	59
Score:	241.00	Matches:	73
Percent Similarity:	47.40%	Conservative:	18
Best Local Similarity:	38.02%	Mismatches:	35
Query Match:	14.10%	Indels:	66
DB:	13	Gaps:	10

US-09-864-291-5 (1-313) X US-10-027-632-290832 (1-593)

```

QY 45 G1YThirValysG1YThirLeuPheIleuThrSerTyrArgValValPheAlThrSerHis 644
Db 48 GGGCAAGAAAGGCACCATCTACCTTACCCCTTTG-----ATCTTTCTGTGCAG-- 988
QY 65 LeuValAlaIlePheMetLeuSerPheMetMetProPheGlyLeuMetSerAlaPyrThr 844

```

```

Db      99  ---GCGAAGAGATGCAAGCAGTCTCTGTAATGTCATTATCTTGAAGGCTTATGAG 155
Qy      85  IIEGUGLNPProIIePheAlaProAntyTrieLygLyThrlIeGlnAlaIaProGly 104
Db      156  ATCAAGCAGCCTGCTGTTTGAACAACATGATCAAGGAAACAGTGAACCGAAGCAGGC 215
Qy      105  GLYGlyTTPGlyUGlyGlnAlaValPheLyLeuSerPheAryLygLyGlyAlaIleGlu 124
Db      216  GGTGCTGGAGAGGCTCTGCTTCTGGCAAGTCCGCTTCACGGCAGAGGGGCGCACTGAA 275
Qy      125  PheAlaGlnLeuMetValIyAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 144
Db      276  TCTGACAGCAGAGTCTCCAGGTGATCATCTCAAGCGTCCAGGGGTGAAGCCCTC----- 329
Qy      145  ValAntyTTPheAapThrSerGlyLeuTrieIleThr---ValProGlyAlaIaIa 163
Db      330  -----AGTGAAGCTACGCGCACTCTCAAGTCTGCTGCTGCAAGACA 368
Qy      164  -----ValCySerSerGlnThrProCy---ProIaTyProIle 176
Db      369  TGTGCTTTCCCGCGCAGTCACTAGTGAATGTATCTGCTGCTGCTGCTGCTGCTGCTGCT 425
Qy      177  ValIleTyGlyProProProGlyTyThrValGlnProGlyGlyTyGlyThrPro 196
Db      426  -----TCTCAACCACTTCCCGTGAAGTCTTATCCAGGA-----CCT 461
Qy      197  ProGlyUGlyTyGlyAlaGlnProGlyTyGlyTyGlyAlaProPro-----Met 212
Db      462  CCCAAGATGACAGGAGCCATG-----GGGTATGTGCAACCCCGCACCTGCTGCTGCT 515
Qy      213  GLYTyGlyAlaProProValGlyTyGlyValPro 224
Db      516  GGGCCCATGGAACCTCCGCTGACGCCCAAGATGTTCC 551

```

RESULT 8

```

US-10-027-632-290832
/ Sequence 290832, Application US/10027632
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ POLYMORPHISMS IN THE HUMAN GENOME
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 290832
/ LENGTH: 593
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-290832

```

Alignment Scores:

```

Pred. No.: 1,68e-15 Length: 593
Score: 241.00 Matches: 73
Percent Similarity: 47.40% Conservative: 18
Best Local Similarity: 38.02% Mismatches: 65
Query Match: 14.10% Indels: 36
DB: 14 Gaps: 10

```

US-09-864-291-5 (1-313) x US-10-027-632-290832 (1-593)

```

Qy      45  GLYThryLyGlyGlyThryLeuPheLeuThrSerTyArgValIaPheValThrSerHis 64
Db      48  GGGACCAAGAAAGGACCACTTACCTTACCCCTTC-----ATCTTCTGTCAGAG--- 98
Qy      65  LeuValAntyAPProMetLeuSerPheMetMetProPheGlyLeuMetSerApyThr 84
Db      99  ---GCGAAGATGCCACGCAAGTCTCTGTAATGTCATTATCTTGAAGGCTTATGAG 155
Qy      85  IIEGUGLNPProIIePheAlaProAntyTrieLygLyThrlIeGlnAlaIaProGly 104
Db      156  ATCAAGCAGCCTGCTGTTTGAACAACATGATCAAGGAAACAGTGAACCGAAGCAGGC 215
Qy      105  GLYGlyTTPGlyUGlyGlnAlaValPheLyLeuSerPheAryLygLyGlyAlaIleGlu 124
Db      216  GGTGCTGGAGAGGCTCTGCTTCTGGCAAGTCCGCTTCACGGCAGAGGGGCGCACTGAA 275
Qy      125  PheAlaGlnLeuMetValIyAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 144
Db      276  TCTGACAGCAGAGTCTCCAGGTGATCATCTCAAGCGTCCAGGGGTGAAGCCCTC----- 329
Qy      145  ValAntyTTPheAapThrSerGlyLeuTrieIleThr---ValProGlyAlaIaIa 163
Db      330  -----AGTGAAGCTACGCGCACTCTCAAGTCTGCTGCTGCAAGACA 368
Qy      164  -----ValCySerSerGlnThrProCy---ProIaTyProIle 176
Db      369  TGTGCTTTCCCGCGCAGTCACTAGTGAATGTATCTGCTGCTGCTGCTGCTGCTGCTGCT 425
Qy      177  ValIleTyGlyProProProGlyTyThrValGlnProGlyGlyTyGlyThrPro 196
Db      426  -----TCTCAACCACTTCCCGTGAAGTCTTATCCAGGA-----CCT 461
Qy      197  ProGlyUGlyTyGlyAlaGlnProGlyTyGlyTyGlyAlaProPro-----Met 212
Db      462  CCCAAGATGACAGGAGCCATG-----GGGTATGTGCAACCCCGCACCTGCTGCTGCT 515
Qy      213  GLYTyGlyAlaProProValGlyTyGlyValPro 224
Db      516  GGGCCCATGGAACCTCCGCTGACGCCCAAGATGTTCC 551

```

RESULT 9

```

US-10-029-386-24981
/ Sequence 24981, Application US/10029386
/ Publication No. US20030194704A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ EXPRESSION ANALYSIS TWO
/ FILE REFERENCE: ABOWICA-X-2
/ CURRENT APPLICATION NUMBER: US/10/029,386
/ CURRENT FILING DATE: 2001-12-20
/ NUMBER OF SEQ ID NOS: 34288
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 24981
/ LENGTH: 511
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ OTHER INFORMATION: MAP TO 297205.1
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
US-10-029-386-24981

```

Alignment Scores:

```

Pred. No.: 2,23e-15 Length: 511
Score: 239.00 Matches: 51
Percent Similarity: 45.97% Conservative: 6
Best Local Similarity: 41.13% Mismatches: 59
Query Match: 13.98% Indels: 8

```


US-10-007-926A-215

Alignment Scores:	
Pred. No.:	4,73e-12
Score:	214.00
Percent Similarity:	47.37%
Best Local Similarity:	42.76%
Query Match:	12,52x
DB:	13
	13
Length:	2176
Matches:	65
Conservative:	7
Mismatches:	58
Indels:	22
Gaps:	11

US-09-864-291-5 (1-313) X US-10-007-926A-215 (1-2176)

[illegible]

RESULT 14

```

US-09-918-995-4554
; Sequence 4554, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4554
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-4554

```

Alignment Scores:	
Pred. No.:	1.3e-12
Score:	50.77%
Percent Similarity:	51.77%
Best Local Similarity:	46.15%
Query Match:	12.35%
DB:	11
length:	433
Matches:	60
Conservative:	6
Mismatches:	44
Gaps:	10
Indels:	20

US-09-864-291-5 (1-313) X US-09-918-995-4554 (1-439)

```

Dy      172  ProLalatyProIleValIleTyTgIyProProProProIyTyTnValaGlnProGly 1910
Db      49  CCAAGGCTATCCCCCAACAAGGCTAC---CCACTTTCCCTGATATCTCTCCGACAGTCAg 1050
Dy      192  GluTyTgIyThnProProGluGlyTyTgIyAlaGlnProGlyTyTgIyAlaProPro 2110
Db      106  GAGTCATCTTTTCCCCCTCTGAGTCAAGATCTTATCTCAAGTGGCTTT-----CTCCCA 1590
Dy      212  MetGlyTyTgIyAla-----ProProValaGlyTyTgIyValaProProGlyGly 2270
Db      160  ATGGAGAGAGGTGCTTACCCACCAAGTGTCCAAAGTAGTGCTAC-----CCAGAGGCT 2100
Dy      228  TyGlyValaPro---ProGlyTyTgIyValaProProGlyGlyTyTgIyAlaPro--- 2450
Db      211  GAGAGCTACCTGCGCTGAGGTTATCTCAAGCCCTGAGGCTATCTCGTGTGCCCAAG 2700
Dy      246  ProGlyGlyTyT---GlyValaProProGly---GlyTyTgIyAlaProPro 2660
Db      271  CCAAGGGGAGCTCATCTTATCCCGAGGTCTCCCAAGGCCAAGATATGGAGTCCCAcCA 3300
Dy      261  GllyGlyTyTgIyAlaProProLalatyTyTgIyAlaProProLalaglyAenGluAlaLeu 2800
Db      331  GGTGGAGCAAGGCTTT-----TCGGGATATCAcCAAGCACTTCAGATCTTATGAGGT 3840
Dy      281  ProProLalatyTgIuAlaProseRalagly 290
Db      385  GGTTCAGCA---CAGGTTCACTACTGCT 411

```

RESULT 15

```

08-10-156-761-3543/c
? Sequence 3543, Application US/10156761
? Publication No. US20030119018A1
? GENERAL INFORMATION:
? APPLICANT: OMURA, SATOSHI
? APPLICANT: IKEDA, HARO
? APPLICANT: ISHIKAWA, JUN
? APPLICANT: HIRAKAWA, HIROSHI
? APPLICANT: SHIBA, TADAYOSHI
? APPLICANT: SAKAKI, YOSHIYUKI
? APPLICANT: HATTORI, MASAHIRA
? TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
? FILE REFERENCE: 249-262
? CURRENT APPLICATION NUMBER: US/10/156,761
? CORRECT FILING DATE: 2002-05-29
? PRIOR APPLICATION NUMBER: JP 2001-204089
? PRIOR FILING DATE: 2001-05-30
? PRIOR APPLICATION NUMBER: JP 2001-272697
? PRIOR FILING DATE: 2001-08-02
? NUMBER OF SEQ ID NOS: 15109
? SEQ ID NO 3543
? LENGTH: 2709
? TYPE: DNA
? ORGANISM: Streptomyces avermitilis
? FEATURES:
? NAME/KEY: CDS
? LOCATION: (1)..(2709)
US-10-156-761-3543

```

Alignment Scores:	
Pred. No.:	1,78e-11
Score:	209.50
Percent Similarity:	36.24%
Best Local Similarity:	29.87%
Query Match:	12.26%
DB:	15
Gaps:	16
Length:	270e
Matches:	89
Conservative:	19
Mismatches:	81
Indels:	110
Gaps:	16

US-09-864-291-5 (1-313) X US-10-156-761-3543 (1-2709)

QY		97	GlyThrIleGlnAlaA1aProGlyValGlyTyrPoiuglGlnAlaValPheIylLeuSer	116
			: : : :	
Db		1786	GggGgcGgTAAGCGACGGCTGTAGCCGGACCAcGAcCGATCGACTGCACCGATTGGCG	1722

```
Oy 117 ---PheArgLySGlyValAlaIleGluPheAlaGluLeuMetVal-----Lys 131
Db 1726 GATACCCGTTACCCCGGGGCTGCTGACCGGGCATCGCGCGCATCGGCTGACCGG 1667
Oy 132 AlaIaSerAlaAlaAlaArgGlyIleProLeuGlySerValAsnTyrrPheAspThr 151
Db 1666 GCATCGGCTGGGGCGCCCATGGCTGTCCCGCTGGGGCGGGCCAGGCC--TGGCGACA 1608
Oy 152 SerGlyLeuTyrrIleIleThrValProGlyAlaAlaValCysSerSerGlnThrProCys 171
Db 1607 CCGGAGGCGTACCGGCAAGGCAATGCCCGGAGCA-----CCGGAGCACCCCGGC 1560
Oy 172 ProAlaTyrrProIleValIleTyrrGlyProPro----- 182
Db 1559 GGGGGCAAGACC-----GACCGCCGACCGGGGAGACCGGCGACATCGTCTCG 1512
Oy 183 -----ProProGlyTyrrThrValGlnProGlyGluTyrrGlyThrProProGlu 198
Db 1511 GCGTGTGACACACACCGGGGGCCACCGGGGGCTCCAGAGAGACCGGGTGGCGCC----- 1458
Oy 199 GlyTyrrGlyAlaGlnProGlyGlyTyrrGlyAlaProPomMetGlyTyrrGlyAlaPro--- 217
Db 1457 ---GAGAGCATGCCCGGGGACCGGGGGCACCCGGCGGGAGAGCGGCGCGCCCGG 1404
Oy 218 -----ProValGlyTyrrGly 222
Db 1403 CCGCCACCGAGGAGCCGGCGAGCATGTGTGCGCGGTATGATGACGCGACCGGGA--GCG 1347
Oy 223 ValProProGlyGlyTyrrGlyValPro----- 231
Db 1346 GTACCGCCCGGAGTTCGGGGCACCGGGCGCGCGGAGACACCGGCTGACCGGGAGCA 1287
Oy 232 -----ProGlyGlyTyrrGlyValPro-----ProGly 240
Db 1286 CCGGGCGGGCCCGGAGGTTCCGGAGCGCCCGAGCGCCGGAGCGCGGCGGAGC 1227
Oy 241 GlyTyrrGlyAlaPro-----ProGlyGlyTyrrGly----- 250
Db 1226 GCGTGGGGCGCGCCGGGCGCGGACCGGGCGCGCGCGGGGTCCGGAGACATCGTGGCCGCG 1167
Oy 251 -----ValProProGlyGlyTyrrGlyAlaProProGlyGlyTyrrGlyAlaProProAla 268
Db 1166 TGATGGACACCCCGGGC-----GCGTACCGCCGAGAGGTTCCGCGTACCGGGCGCG 1113
Oy 269 -----GlyTyrrGlyAla----- 272
Db 1112 CTCGGCGGATTCGGGCGCGGAGCGCTCGGGCCCGTCCGGACCGAGTTGGAGAGCGAGC 1053
Oy 273 -----ProProAlaGlyAsnGluAlaLeuProProAlaTyrrGluAlaPro 287
Db 1052 TCGGTGGACATACCCCGCGCGGCG-----GTACCCGAGACACCGGGCGCG 1005
Oy 288 SerAlaGlyAsnThrAla-AlaSerHisArgSerMetThrAlaGlnGlnGlu 304
Db 1004 GACGCGGAGGATGGACACCGGGCGCACCCGCGTTCCAGGGGACACCGGAA 953
```

Search completed: December 16, 2003, 20:57:12
Job time : 337.646 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 17:31:55 | Search time 2021.93 Seconds
(without alignments)
3762.388 Million cell updates/sec

Title: US-09-864-291-5
Perfect score: 1709
Sequence: 1 MAVNQSHTSRGALIPSGR.....ASHSRMTAQGTSLPTSSS 313

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USFO_spool_p/US09864291/runat.15122003.160859.23318/app.query.fasta.1.846
-DB=BST -QFMT=faa+ap -SUPFIX=ret -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HRAPIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09864291.QCEN.1.1.3724 @runat.15122003.160859.23318 -NCPU=6 -ICPU=3
-NO MAP -LARGESUBJECT -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXt=7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_huv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vir:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	983	57.5	1079	12	BM564167
2	935.5	54.7	1414	11	AK015863
3	730.5	42.7	717	14	BY715414
4	653.5	38.2	775	10	BQ722815
5	616	36.0	663	10	BG699398
6	578.5	33.9	932	13	BU107273
7	552	32.3	868	14	CD255288
8	552	32.3	931	13	BU914288
9	540.5	31.6	723	9	AJ452591
10	539.5	31.6	700	13	BU284977
11	536	31.4	636	12	BM490729
12	536	31.4	862	13	BQ735286
13	528.5	30.9	892	14	CD301781
14	517	30.3	719	14	CB593075
15	513	30.0	655	9	AL898723
16	513	30.0	661	9	AL972453
17	513	30.0	663	9	AL896682
18	506	29.6	647	9	AL849651
19	505	29.5	663	9	AL881810
20	504	29.5	816	13	BU916890
21	500	29.3	568	12	B1390687
22	495	29.0	607	12	BM190975
23	483	28.3	560	13	BU103740
24	481.5	28.1	1165	12	BM923449
25	480.5	28.1	594	12	BM190972
26	480	28.1	735	14	CA364472
27	478.5	28.0	635	9	AL889954
28	476	27.9	613	9	AL889954
29	475	27.8	611	9	AL849211
30	475	27.8	974	13	BQ876783
31	475	27.8	1027	12	BM921835
32	474.5	27.8	970	12	BM424229
33	474.5	27.8	992	12	BM917483
34	472.5	27.6	1024	12	BM451662
35	471.5	27.6	948	13	BU848882
36	471	27.6	865	13	BQ427510
37	471	27.6	1018	12	BM470322
38	470.5	27.5	1068	13	BQ277733
39	470	27.5	914	13	BQ710173
40	469.5	27.5	896	13	BQ212538
41	469.5	27.5	951	14	CD515402
42	468.5	27.4	920	13	BU165608
43	468.5	27.4	1013	12	BM489934
44	468.5	27.4	1042	12	BM479536
45	468	27.4	726	14	CB591559

ALIGNMENTS

RESULT 1
LOCUS BM564167
DEFINITION AGENCOURT_6560149 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742129
5', mRNA sequence.
ACCESSION BM564167
VERSION BM564167.1 GI:198811738
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1079)

Alignment Scores:

Pred. No.:	6,896-42	Length:	717
Score:	730.50	Matches:	145
Percent Similarity:	64.45%	Conservative:	20
Best Local Similarity:	56.64%	Mismatches:	54
Query Match:	42.74%	Indels:	37
DB:	14	Gaps:	4

US-09-864-291-5 (1-313) x BY715414 (1-717)

```

Qy 1 MetAlaValaAngInSerH1eThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
Db 55 ATGCAAGTGAACAGAACCATACAGTGAACGCGCTGGCGGCATCCCTCATGGTGA 114
Qy 21 SerValLeuYsgInCyGluAspValAspLeuCySphLeuGlnYsProValGluSer 40
Db 115 AGTCTCTTGAAGAATGCTTCAAGTGAACCTCTCCCTCCCGAGTCAACCCGAGCTCC 174
Qy 41 TytleupheAnglyThrIlySgIyThrIleupheLeuThrSerTyTzArgValIlePhe 60
Db 175 AACCTCTTGAAGTGAAGAAGAGGAGGAGCTCTGTTCTCAGCTCATACCGAGTATTT 234
Qy 61 ValThrSerH1sLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
Db 235 GTGACTTCGCGTGAACAGACATCCATGTTTCTTTACAGATCCATTCATCATATG 294
Qy 81 SerAspCyThrIleGluGlnProIlePheAlaProAsnTyTleYsgIyThrIleGln 100
Db 295 AATAAGTCACTGTGGAACAACCATCTTGTGCAAACTCATTTAAAGGAGCAATTCAG 354
Qy 101 AlaAlaProGlyGlyIyTPGlyGlyGlnAlaValaPheYsLeuSerPheArgIyGly 120
Db 355 GCAGCTCAGAGCGGTGGTGGAAAGTTCGTCTTAAATAAGTCTTCCGAAAGGA 414
Qy 121 G1ValaIleGluPheAlaGlnLeuMetValIyAlaAlaSerAlaAlaAlaArgIyIle 140
Db 415 GGTCTATTCAGCTTGTCCAGTGTGATGCGCAAGCTCTGCGCTGCCAGAGATT 474
Qy 141 ProLeuGlySerValAsnTyTTPheAspThrSerGlyLeuTyTleIleThrValPro 160
Db 475 CCATCAGAGTGAAGTGAAGTTCGATCGAGCCCTCAGAAATTTATGATCATCAT- 528
Qy 161 G1ValaIleValaCySerSerGlnThrProCyProAlaTyTTProIleValIleTyGly 180
Db 528 ----- 528
Qy 181 ProProProGlyTyTThrValaGlnProGlyGlyTyTThr---ProProGlyGly 199
Db 529 -----GGGACAGAAACATGTATGCCCCACAGCA 558
Qy 200 TyTgIyAlaGlnProGlyGlyTyTgIyAlaProProMetGlyTyTgIyAlaProProVal 219
Db 559 TACCAAGTT-----GCCTATGAGAGCCCACTCGGAGATATGAGGAGCTCAACCGTG 609
Qy 220 G1TyTgIyAlaProProGlyGlyTyTgIyAlaProProGlyGlyTyTgIyAlaProPro 235
Db 610 GGAATACGAGTCCATCTGACAGGATATGAGACCCCACTCGGAGTATGAGAGCCGCC 669
Qy 240 G1YgIyTyTgIyAlaProProGlyGlyTyTgIyAla---ProProGly 254
Db 670 GTGGATATGTAACCCCATCTCCGAGATGATGCTGCTCTCGGA 717

```

RESULT 4

BG722815

LOCUS BG722815 775 bp mRNA linear EST 08-MAY-2001

DEFINITION 602695278P1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4827427 5',

mRNA sequence.

ACCESSION BG722815

VERSION BG722815.1 GI:14002002

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
1 (Pages 1 to 775)
Mammalia; Rutheria; Primates; Catarrhini; Homnidae; Homo.
NITH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
JOURNAL
COMMENT
Contact: Robert Strauberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palokovics, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
<http://image.llnl.gov>
Plate: LIML0743 row: 1 column: 20
High quality sequence stop: 651.
Location/Qualifiers
1. 775

FEATURES

source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4827427"
/lab_host="DH10B"
/clone_11b="NIH MGC 97"
/note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site: 1: BamHI, Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TGTGTTTGTGTTTGTG-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGR), National
Institutes of Health). Note: this is a NIH_MGC library."

```

BASE COUNT

191 a 212 c 193 g 179 t

ORIGIN

Alignment Scores:

Pred. No.:	1,876-36	Length:	775
Score:	653.50	Matches:	144
Percent Similarity:	71.37%	Conservative:	28
Best Local Similarity:	59.75%	Mismatches:	48
Query Match:	38.24%	Indels:	23
DB:	10	Gaps:	4

US-09-864-291-5 (1-313) x BG722815 (1-775)

```

Qy 1 MetAlaValaAngInSerH1eThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
Db 54 ATGCAAGTGAACAGAACCATACAGTGAACGCGCTGGCGGCATCCCTCATGGTGA 113
Qy 21 SerValLeuYsgInCyGluAspValAspLeuCySphLeuGlnYsProValGluSer 40
Db 114 AGTCTCTTGAAGAATGCTTCAAGTGAACCTCTCCCTCCAGATCAGAAAGCTCA 173
Qy 41 TytleupheAnglyThrIlySgIyThrIleupheLeuThrSerTyTzArgValIlePhe 60
Db 175 AACCTCTTGAAGTGAAGAAGAGGAGGAGCTCTGTTCTCAGCTCATACCGAGTATTT 233
Qy 61 ValThrSerH1sLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
Db 234 ATAAGTCAATGCTCCATCAGTGAATCCCATGTTTCTTTATGATGATGATGATGATG 293
Qy 81 SerAspCyThrIleGluGlnProIlePheAlaProAsnTyTleYsgIyThrIleGln 100
Db 294 ACAAGCTCACTGTGGAACAACGATTTGTCTGCAAACTCATTTAAGGAACTATTGAG 353
Qy 101 AlaAlaProGlyGlyIyTPGlyGlyGlnAlaValaPheYsLeuSerPheArgIyGly 120
Db 354 GCAGCTCCATATGATGCTGGAGAGCAAGTACTTAAATTAATGATCTTCAGAAATGGA 413
Qy 121 G1ValaIleGluPheAlaGlnLeuMetValIyAlaAlaSerAlaAlaAlaArgIyIle 140

```

Db 414 GGTGCCATTGAATTGCCAGTTATGTTGAAGTCCCTCTGCTGCTGCCGAGAAATT 473

Qy 141 ProLeuGlySerValAsn-TyTrpPheAspThrSerGlyLeuTyIleIleThValPr 160

Db 474 CCACCTAGAACCTTAACGACTGGTTCAGCTCTATGGGAATTATGTAATTACGGGGA 533

Qy 160 GGLYAlaValCysSerSerGlnThrProCysProAlaTyIleProIleValIleTyG 180

Db 534 AGGGC--AATATGTCACCTCCACAGATGCTTGT-----TCAGTTATGCTATAG 581

Qy 180 TyProProProProGly-TyTrpThValGlnProGlyGluTyIleThProProGluGly 200

Db 562 ACCCCCACTGACGAGATAT-----GAGCCCACTCCCGGAT 620

Qy 200 YTGlyAlaGlnProGlyGlyTyGlyAlaProProMetGlyTyTyGlyAlaProProValG 220

Db 621 ACGGAGCCCACT-GCAGGATATGAGCCCAACCCGTAGAA-GCAGCGCCGCTGCTG 678

Qy 220 TyTyT-----GlyValProProGlyGlyTyTyGlyValPro 231

Db 679 GCATACAGAGCTCAACCTGCTGCAGCAATGGAAGCCCACTCTCGGCAATACGAGAGCC 737

RESULT 5

LOCUS BG699398 663 bp mRNA linear EST 07-MAY-2001

DEFINITION 602679081P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4811804 5', mRNA sequence.

ACCESSION BG699398

VERSION BG699398.1 GI:13967653

KEYWORDS EST.

ORGANISM Homo sapiens (human)

SOURCE Homo sapiens

REFERENCE 1 (bases 1 to 663)

AUTHORS NIH-MGC <http://mgs.mcl.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue procurement: Miklos Palokovits, M.D., Ph.D.

cDNA library preparation: Michael J. Brownstein (NHGRI), Shitaki Toshitaki and Piero Carninci (RIKEN)

cDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLNL0703 row: a column: 21

High quality sequence stop: 660.

Location/Qualifiers

1. 663

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4811804"

/issue_type="hnpocampus"

/lab_host="DH10B"

/clone_11b="NIH_MGC_95"

/note="Organ: Brain; Vector: pBluescriptPCr (modified pBluescript KS+); Site 1: BamHI, Site 2: SalI-XhoI (gtcagag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3', size selected for average insert size 2.5 kb and normalized to RGT 5. This is a primary library enriched for full-length clones and constructed using the cap-trapped method (Carninci, in preparation). Library constructed by M. Brownstein (NHGRI/NHGI, National Institutes of Health). Note: this is a NIH-MGC library."

BASE COUNT 168 a 158 c 169 g 164 t 4 others

ORIGIN

Alignment Scores: 6.71e-34 Length: 663

Pred. No.:

Score: 616.00 Matches: 134

Percent Similarity: 74.09% Conserved: 29

Best Local Similarity: 60.91% Mismatches: 46

Query Match: 36.04% Indels: 14

DB: 10 Gaps: 3

US-09-864-291-5 (1-313) x BG699398 (1-663)

Qy 1 MetAlaValangInserHlethrgIuseArgrglYAlaLeuIleProserGlyGlu 20

Db 37 ATGGCGGGATGATGAGACGACACAGAGAACCGCGGAGCCCTCATCCCTTAAGGGTAA 96

Qy 21 SerValLeuGlyGlnCysGluAspValAspLeuCySphLeuGlnIleProValGluSer 40

Db 97 AGTCTCTTGAAGCCGCTCTCCAGATGTGAGAGCTCTCTCCACAGCCATAGAGAGCTCA 156

Qy 41 TyrlleuPheangIlyThrlYelyGlyThrlleuPheuthrserTyIleValIlePhe 60

Db 157 AATGCTTATGTGTAGAGAAAGACAGAACATG-TTCTCACTTATACCGGGATATTTC 215

Qy 61 ValThrSerHleValAsnAspProMetLeuSerPheMetMetPropheGlyLeuMet 80

Db 216 ATAACTTCATGCTCATCATGATCCATGATGCTTTATATGATGCAATTGATCATG 275

Qy 81 SerAspCythrIleGluGlnProIlePheAlaProAlaTyIleGlyThrlleGln 100

Db 276 ACGAAGCTCACTGTGAAACCAACAGATTGCTCAAACTTATTAAGGAACTATTCAG 335

Qy 101 AlaAlaProGlyGlyTyTrpGluGlyGlyAlaValPheLeuSerPheArglyGly 120

Db 336 GCAAGCTCATATGTGTGTGGAGAGACAGACTTCTTTAATATGATCTTCAAGAAATGGA 395

Qy 121 GlyAlaIleGluPheAlaGlnLeuMetValIleAlaIleAlaIleAlaArglyIle 140

Db 396 GATGCATTAATGATTTGGCCAGTTGATGTGAAAGACCTCTCTGCTGCCGAGAAATT 455

Qy 141 ProLeuGlySerValAsnTyTrpPheAspThrSerGlyLeuTyIleIleThValPro 160

Db 456 CCACCTAGAACCTTAATATGATGCTGTGAGCTCTATGGAATATGTAATTACGGGGA 515

Qy 161 GlyAlaValCysSerSerGlnThrProCysProAlaTyIleProIleValIleTyGly 180

Db 516 GGG--AATATGTCACCTCCACAGATGCTGT-----TCAGTTATGCTATAGGA 563

Qy 181 ProProProProGlyTyTrpThValGlnProGlyGluTyIleThProProGluGlyTy 200

Db 564 GCCCACTGAGAG-----TATGAGCCCACT-CCCGGATAC 601

Qy 201 GlyAlaGlnProGlyGlyTyGlyAlaProProMetGlyTyTyGlyAlaProProValGly 220

Db 602 GAGAGCCC-ACATGAGATATGAGAGCCCAACCGTAGAATGAAGCGCGCTGTGGGA 660

RESULT 6

LOCUS BU107273 932 bp mRNA linear EST 25-NOV-2002

DEFINITION 603111954P1 CSBCHU12 Gallus gallus cDNA clone CHEST60K13 5', mRNA sequence.

ACCESSION BU107273

VERSION BU107273.1 GI:25309754

KEYWORDS EST.

ORGANISM Gallus gallus (chicken)

SOURCE Gallus gallus

REFERENCE Boardman, P.B., Sanz-Bazquero, J., Overton, I.M., Burt, D.W., Bosch, E., Pong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J. A Comprehensive Collection of Chicken cDNAs Curr. Biol. 12 (22), 1965-1969 (2002)

JOURNAL MEDLINE

PMID 12445392

COMMENT Contact: Simon Hubbard

Average insert size 1.5 kb. Constructed by Life Technologies. Note: This is a Xenopus Gene Collection (XGC) library.

BASE COUNT 241 a 244 c 189 g 193 t 1 others

Alignment Scores:

Pred. No.: 2,756-29 Length: 868
Score: 552.00 Matches: 131
Percent Similarity: 53.44% Conservative: 32
Best Local Similarity: 42.95% Mismatches: 88
Query Match: 32.30% Indels: 54
DB: 14 Gaps: 13

US-09-864-291-5 (1-313) x CD255258 (1-868)

1 MetAlaValaAngInSerHleThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
40 ATGTCACTTAATTAAGAACCACTCAAGGGG--GCCGGATCATGTTCAGCAAGGGAG 96
21 SerValleuLeuGInCyseGluAspValaAspLeuCySpheLeuGlnAspProValaGluSer 40
97 AGTATTCTCAGACAGTGAAGATGTGGAACCTCTCTCAGGATATGTCTCAGCAATCA 156
41 TyrLeuPheAenGlyThrlYelYsglyThrlLeuPheLeuThrSerTyrArgValaPhe 60
157 GAAGCATTCAAAGGACAAAGAAAGAGATGCTATATCTCACCGCTACAGGGTGAATCTT 216
61 ValThrSerHleLeuValaAspAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
217 CTAAACCAAGG-----AAAGACCCCATGATGCTCTTCAATATGCAATTCATCACTGATG 270
81 SerAspCyThrIleGluInProIlePheAlaProAsnTyrIleYsglyThrlleGln 100
271 AAAGATGCTCATTAAGAACCACTCTCTCTGCAACTACATCAAGGAAACCATCACT 330
101 AlaAlaProGlyYglYglYtrpGluGlnAlaValaPheYleuSerPheArgYsgly 120
331 GCTGAACCTGAGATGGCTGGAGGGTAAACCTCTTAACTTAATTAACTTAACAGTGA 390
121 GlyAlaIleGluPheAlaGlnLeuMetValaValaAlaSerAlaAlaAlaArgGlyLe 140
391 GGAGCATTAATTAAGAACCAATGATGTTCAAAAGGCACTGCTCCAGACCCCT 450
141 ProLeuGlySerValaAsnTyrTrpPheAspThrSerGlyLeuTyrIleThrValPro 160
451 CCT-----GTTCT 459
161 GlyAlaAlaValaCyseSerSerGlnThrProCySpProAlaTyrProIleValIleTyrGly 180
460 CATGCTGCATAT-----GATATACACTCTCTGCTGAGGA-----GGA 495
181 ProProProProGlyYtrYtrYtrValGlnProGlyYglYtrYtrYtrProProGluGlyYr 200
496 TATGACCAAGGGGGGATACCTCCAGCTCCAGGAATATATACACCCCAACCCCAACATCA 555
201 GlyAlaGlnProGlyYglYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtr 219
556 GGCCCTTATCA-----TATGACCCCTCTGCTATGATGATGATGATGATGATGATGAT 609
220 GlyTyrGlyValaProProGlyYglYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtr 239
610 GCTAAGGATACCG-----TATGCTCCGCAACGAG-----CCT 645
240 GlyGlyYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtr 259
646 GGCATGATCAACCCCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 693
260 ProGlyYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtrYtr 279
694 CAACCACTTATCTCCGACCCCG-----TATATGACCCCAATATATGATGATGATGAT 741
280 LeuProProAlaTyrGlyAlaProSerAlaGly-----AsnThrAla 293

DB 742 ---ACCCAGCACTTCAAGCTCTCTGATTCATGATATCTGTATGCTTGAGAGCAG 798
QY 294 AlaSerHleArgSer 298
DB 799 CAAGCCCAAGAGC 813

RESULT 8
BU914288 931 bp mRNA linear EST 17-OCT-2002
LOCUS AGENCOURT 10492155 NICHD XGC OOI Xenopus laevis cDNA clone
DEFINITION IMAGE:6640777 5', mRNA sequence.
ACCESSION BU914288 1 GI:24096202
VERSION BU914288.1
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis

REFERENCE 1 (bases 1 to 931)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished
COMMENT Contact: Robert Strauberg, Ph.D.
Email: cgapdb-remail.nih.gov
Tissue Procurement: Marthe Rebber, Steven L. Klein, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>
Plate: LIML4223 row: 1 column: 01
High quality sequence stop: 654.

FEATURES

source

1..931
Location/Qualifiers

/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:6640777"
/tissue_type="oocytes"
/lab_host="DH10B (phage-resistant)"
/clone_id="NICHD XGC OOI"
/note="Vector: pCMV-SPORT6, Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by Life Technologies."

BASE COUNT 254 a 256 c 201 g 220 t
ORIGIN

Alignment Scores:

Pred. No.: 2,966-29 Length: 931
Score: 552.00 Matches: 133
Percent Similarity: 55.22% Conservative: 31
Best Local Similarity: 44.78% Mismatches: 91
Query Match: 32.30% Indels: 42
DB: 13 Gaps: 12

US-09-864-291-5 (1-313) x BU914288 (1-931)

1 MetAlaValaAngInSerHleThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
62 ATGTCACTTAATTAAGAACCACTCAAGGGG--GCCGGATCATGTTCAGCAAGGGAG 116
21 SerValleuLeuGInCyseGluAspValaAspLeuCySpheLeuGlnAspProValaGluSer 40
119 AGTATTCTCAGACAGTGAAGATGTGGAACCTCTCTCAGGATATGTCTCAGCAATCA 178
41 TyrLeuPheAenGlyThrlYelYsglyThrlLeuPheLeuThrSerTyrArgValaPhe 60
179 GAAGCATTCAGAGGACAAAGAAAGAGATGCTATATCTCACCGCTACAGGGTGAATCTT 238


```

Qy 61 ValThrSerHisLeuValAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
Db 239 CTAAACAGGG-----AAGATCCCATATGCTTTAATATGCAATTTACATATG 292
Qy 81 SerAspCysThrIleGluGlnProIlePheAlaProAntyTilleyGlyThrIleGln 100
Db 293 AAGAGATGCTTCATAGAACAGCCAGGTCTCTCCCAATCATCAAGGAAACCATAGT 352
Qy 101 AlAlaAlaProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 120
Db 353 GCTGAACCTGAGAGGTGCTGGAGGCTCAAGCTCTTTTAAATTAATTAATTAATGAGA 412
Qy 121 GYAlAlaIleGluPheAlaGlnLeuMetValIleValAlaIleSerAlaAlaAlaArgGlyIle 140
Db 413 GGTCTATTGAATTTGACAAATCATGTCTAAATGACCACTTGCTTCCAGAACCCCT 472
Qy 141 ProLeuGlySerValAsnYrTrpPheAspThrSerGlyLeuYrIleIleThrValPro 160
Db 473 CCT-----GTTCT 481
Qy 161 GYAlAlaAlaValCysSerSerGlnThrProCysProAlaIleYrProIleValIleYrGly 180
Db 482 CATGCTGCATAT-----GGATACACACCTGCTCCAGAA-----GGC 517
Qy 181 ProProProProGlyYrThrValGlyProGlyGlyYrThrProProGlyGlyYr 200
Db 518 TATGCAATAGCGGGTACCTCCAGCTCCAGAAATATATACACCCCAACCATCA 577
Qy 201 GYAlaGlnProGlyGlyYrGlyYrAlaProPheMet---GlyYrGlyAlaPro----- 217
Db 578 GGCCCTTATCCA-----TATGACCCCTGCTATGATGATGATGACCACTCCCAAT 631
Qy 218 ProValGlyYrGly---ValProProGlyGlyYrGlyYrAlaProProGlyYrGly 236
Db 632 CCAATGGAATACCGTATGCTCCACAGAGAA-----CCGCGCATGATACCA 679
Qy 237 ValProProGlyGly-----YrGlyYrAlaProProGlyGlyYrGlyYrAlaPro 253
Db 680 CCCCTCTGAAATGAATCCCATTTACATGAGCTCTCCACACCTTATCTCGAAGCCCC 739
Qy 254 GYAlYrGlyYrAlaProProGlyGlyYrGlyYrAlaProProAlaGlyYrGlyYrAlaPro 273
Db 740 TACAAAT---GGAGCCCTCATAGGAACCCAGCACCTTCTGCTCCCACTGAGATGATG 796
Qy 274 ProAlaGlyAsnGlyAlaLeuProProAlaIleYrGlyAlaProSerAlaGly 290
Db 797 AATGCTGTATGCTGAGAGTGAAGCAAGCTGCAAAAGCTGCATCCAGTGGC 847

RESULT 9
LOCUS AJ452591 723 bp mRNA linear EST 22-APR-2002
DEFINITION AJ452591 riken1 Gallus gallus cDNA clone 31hr1, mRNA sequence.
VERSION AJ452591.1 GI:20262687
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
AUTHORS Buerstedde, J.M.
TITL Gallus gallus Bursal Lymphocyte EST
JOURNAL Unpublished
COMMENT Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinstier. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dc40est.html.
FEATURES
Source
1..723
/organism="Gallus gallus"
/mol_type="mRNA"

```

```

/db_xref="taxon:9031"
/clone="31hr1"
/cell_type="Bursal Lymphocyte"
/dev_stage="2-3 weeks old"
/clone_11b="riken1"
/note="CB inbred strain"
BASE COUNT 161 a 205 c 181 g 172 t 4 others
ORIGIN
Alignment Scores:
Pred. No.: 1,44e-28 Length: 723
Score: 540.50 Matches: 126
Percent Similarity: 52.23% Conservative: 26
Best Local Similarity: 43.30% Mismatches: 66
Query Match: 31.63% Indels: 74
DB: 9 Gaps: 11

US-09-864-291-5 (1-313) x AJ452591 (1-723)
Qy 1 MetAlaValAsnGlnSerHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
Db 12 ATGGCGCTCATATGAGCAACCACTGAAG---GAAGCGGTCTCGTCTCCCAATGCCAG 68
Qy 21 SerValLeuYrGlnCysGluAspValAspLeuCysPheLeuGlnYrProValGlySer 40
Db 69 AGCATTTCAAGCAATGTAAAGATGTGAGCTCTCTTCATGATCAATGACGGCAAGCTT 128
Qy 41 TyrLeuPheAsnGlyThrIleYrGlyGlyThrLeuPheLeuThrSerYrArgValValPhe 60
Db 129 GAAGCCTCAAGAGCAACAAAGAGAAATGCTATCTCAACCCATACAGATATCTTC 188
Qy 61 ValThrSerHisLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
Db 189 GTGTCAAGGCGC-----AAGATCTCATGCTGTCTTTATATATATATATATTTG 242
Qy 81 SerAspCysThrIleGluGlnProIlePheAlaProAntyTilleyGlyThrIleGln 100
Db 243 AAGGCTGCTCTTATGAGAGCTCTTTCTCTGCTAATTAATCAAGAGCAATTCAG 302
Qy 101 AlAlaAlaProGlyGlyYrTrpGlyGlyAlaValAlaPheYrLeuSerPheArgGly 120
Db 303 GCTGAAGCAGAGGTGCTGGAGAGGCGAGGAACCTTTAACTGACTTTCAACAGGGA 362
Qy 121 GYAlAlaIleGluPheAlaGlnLeuMetValIleValAlaIleSerAlaAlaAlaArgGlyIle 140
Db 363 GGAGCATCATGATTTGACAGCTGATGTTCAAAAGCTGCTCTGCTCCCACTGAGATT 422
Qy 141 ProLeuGlySerValAsnYrTrpPheAspThrSerGlyLeuYrIleIleThrValPro 160
Db 423 CTTCTC----- 428
Qy 161 GYAlAlaAlaValCysSerSerGlnThrProCysProAlaIleYrProIleValIleYrGly 180
Db 428 ----- 428
Qy 181 ProProProProGlyYrThrValGlnProGlyGlyYrGlyYrThrProProGlyGlyYr 200
Db 429 ---CAAGCCCTGCTAT----- 449
Qy 201 GYAlaGlnProGlyGlyYrGlyYrAlaProProMetGlyYrGlyYrAlaProProValGly 220
Db 450 ACACTGTACCTGAGAGGTAT---GCACCT-----GTCCCACTGCT--- 488
Qy 221 TyrGlyYrAlaProProGlyGlyYrGlyYrAlaProProGlyGlyYrGlyYrAlaProProGly 240
Db 489 -----CCAAGAGGTATGACCTGCTCCAGAGAGGTATGACCTGCTCCGGGG 536
Qy 241 GYTYrGlyYrAlaProProGlyGlyYrGlyYrAlaProProGlyGlyYr-----YrGly 257
Db 537 GGCTAT---GCTCTCTC-----NCACCACTCCAAATGATCTTATCTTATGCA 584
Qy 258 AlaProPro---GlyYrGlyYrGlyYrAlaProProAlaGlyYrGlyYrAlaProProAlaGly 276

```

Db 585 CCACCTCCGATGATGCTACGACCTGCTCCCAACCCATGAGGAT-CCGTATGCCAG 643
 Qy 277 AAGGUAAlaLeuProProAlaTyrGluAlaPro 287
 Db 644 ATCCAGATATTATTACCCACCCGCTGCCACCA 676
 RESULT 10
 BU284977 700 bp mRNA linear EST 27-NOV-2002
 LOCUS 603603379P1 CSEGCN54 Gallus gallus cDNA clone CHEST581n20 5', mRNA
 DEFINITION
 sequence.
 ACCESSION BU284977 GI:25734433
 VERSION BU284977.1
 KEYWORDS
 SOURCE Gallus gallus (chicken)
 ORGANISM
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 700)
 Boardman, P.E., Sanz-Bzquerro, J., Overton, I.M., Burt, D.M., Bosch, B.,
 Fong, W.T., Tickle, C., Brown, M.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken CDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 JOURNAL PUBMED
 COMMENT 12445392
 Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1..700
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Compton line 15T"
 /db_xref="taxon:9031"
 /clone="CHEST581n20"
 /sex="Female"
 /tissue_type="not cerebrum or cerebellum"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_1lb="CSEGCN54"
 /note="Organ: Brain; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunt-ended, ligated to NotI adapters, digested with EcoRI
 , size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

BASE COUNT 162 a 196 c 178 g 164 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.64e-28 Length: 700
 Score: 539.50 Matches: 124
 Percent Similarity: 51.758 Conservative: 24
 Best Local Similarity: 43.368 Mismatches: 67
 Query Match: 31.578 Indels: 72
 DB: 13 Gaps: 9

US-09-864-291-5 (1-313) X BU284977 (1-700)

Qy 1 MetAlaValAlaGlnSerHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
 Db 3 ATGGGCTCATAGGACCACTCGAAG---GAAAGCGGTGTCGTCGCCACCAAGCCG 59
 Qy 21 SerValLeuGlyGlnGlyGluAlaGlyValAlaSerGlySerPheLeuGlnGlyProValGlnSer 40
 Db 60 AGCATTCCTCAAGCAATGTAAGATGTGAGCTCTCTCTCACTGACATGACGAGGAGCTT 119
 Qy 41 TyrLeuPheAlaGlnGlyThrLeuGlyGlyThrLeuPheLeuThrSerTyrArgValAlaPhe 60
 Db 120 GAGGCTTCGAAAGGACCAAGAGGATGCTGATATCTACACCCATACAGATATCTTC 179
 Qy 61 ValThrSerHisLeuValAlaAspProMetLeuSerPheMetProPheGlyLeuMet 80
 Db 180 GTGTCAAGAGGC-----AAGATCTTATGCTGCTTTATGATGCCGTTTATTATGTTG 233
 Qy 81 SerArgGlyThrIleGlnGlnProIlePheAlaProAlaGlyTyrIleGlyThrIleGln 100
 Db 234 AAGAGTCTCTATGACAGACCTCTTCTCTCTATATTCATCAAGAGGACAGATTCAG 293
 Qy 101 AlaAlaProGlyGlyGlyTyrGlyGlyGlyAlaValPheLeuSerPheArgGlyGly 120
 Db 294 GCTAGAGGACGAGAGGTGCTGGAGAGGCGAGGAGACGTTTAACTGACTTCAACAGCGGA 353
 Qy 121 GlyAlaIleGluPheAlaGlnLeuMetValIleAlaAlaSerAlaAlaAlaArgGlyIle 140
 Db 354 GAGGCAATCGAGTTTGAACAGCTGATGTTCAAGCTGCTCTAGGCTTCCAGTGAAGTT 413
 Qy 141 ProLeuGlySerValLeuTyrTrpPheAspThrSerGlyLeuTyrIleIleThrValPro 160
 Db 414 CCTCTC----- 419
 Qy 161 GlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrProIleValIleTyrGly 180
 Db 419 ----- 419
 Qy 181 ProProProGlyGlyTyrThrValGlnProGlyGlyGlyTyrProProGlyGlyTyr 200
 Db 420 ---GAGCCCTCGGCTAT-----GATATC 440
 Qy 201 GlyAlaGlnProGlyGlyTyrGlyAlaProProMetGlyTyrGlyAlaProProValGly 220
 Db 441 AACCTGTATCTGAGGAGGAT---GCACT-----GTCCACCTGCT--- 479
 Qy 221 TyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProProGly 240
 Db 480 -----CGAGAGGATGACCTGCTCCAGAGGATGACCTGCTCGGAG 527
 Qy 241 GlyTyrGlyAlaProProGlyGlyTyrGlyValProProGlyGlyTyr-----Gly 257
 Db 528 GGCTATGCTCCCTCCGCA-----ACTCAATGATGCTTATCTTATGACCA 577
 Qy 258 AlaProProGlyGlyTyrGlyAlaProProAlaGlyTyrGlyAlaProProAlaGlyAsn 277
 Db 578 GCTCCGATGAATGCTTACGAGACCTGCTCCAGCCCATGAGGATTCGATAGCCAGAT 637
 Qy 278 GluAlaLeuProProAla 283
 Db 638 CCAGGTATGTACCCAGCG 655
 RESULT 11
 BM490729 636 bp mRNA linear EST 07-FEB-2002
 LOCUS BM490729
 DEFINITION psp2n.pk004.e6 Normalized Chicken Pituitary/Hypothalamus/Pineal
 Library (psp2n) Gallus gallus cDNA clone psp2n.pk004.e6 5' similar
 to gb|AAD10951.1 (U79458) WW domain binding protein-2 (Homo sapiens
), mRNA sequence.
 ACCESSION BM490729
 VERSION BM490729.1 GI:18611660
 KEYWORDS
 SOURCE Gallus gallus (chicken)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 636)
Porter, T.B. and Cogburn, L.A.
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
Library, USDA/IRAPs Animal Genome Project
Unpublished
Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickbase.udel.edu.

FEATURES
source
1. 636
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Commercial broiler chickens"
/db_xref="taxon:9031"
/clone="pdp2n.pk004.e6"
/sex="Male and Female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
/dev_stage="Embryonic (d12,d14,d19); post-hatch (1,3,5,7,9 weeks)"
/lab_host="B. coli BMDH108"
/clone_lib="Normalized Chicken
Pituitary/Hypothalamus/Pineal Library (pdp2n)"
/note="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue at different
ages. Single pass sequencing from 5'-end"

BASE COUNT 143 a 174 c 173 g 146 t

ORIGIN

Alignment Scores:
Pred. No.: 2.6e-28 Length: 636
Score: 536.00 Matches: 120
Percent Similarity: 52.00% Conservative: 23
Best Local Similarity: 43.64% Mismatches: 58
Query Match: 31.36% Indels: 74
Gaps: 10

US-09-864-291-5 (1-313) x BM490729 (1-636)

QY 1 MetAlaValaAnglnSerHrthrglUserHrgrglYalaenlleProserGlyglu 20
Db 23 ATGGCGCTCAATAGGCGCACCTCAAG---GAAGCGGTGTCGTGCCCAATAGCGAG 79
QY 21 SerValleuysglnCysgluAspValAspleuCybphleuGlnlyspProValGluSer 40
Db 80 AGCATTTCTAGAGATTAAGATGAGAGCTCTCTTCAGTGCATGACGAGCGAGCCTT 139
QY 41 TyrleuPheAanglyThrlylysglyThrleuPheleuThserTyrtatgValaPhe 60
Db 140 GAGGCGCTTCAAGGCGACCAAGGAGGAGATCTGATCTCAAGCCCATACAGAGATCTTC 139
QY 61 ValThSerHsleuValaAspPromeleuSerPheMetCpProPheGlyleuMet 80
Db 200 GTGTCAAGGGC-----AAGGATCTCATCTGCTTTATGATGACCAATTTATTTGGTG 253
QY 81 SerAspCyethrlleGluGlnProIlePheAlaProAsnTyrtlleysglyThrIleGln 100
Db 254 AAGGGGCTCTATAGAGCGCTCTTTCTCTGCTAATATACCAAGGAGCAAGATTGAG 313
QY 101 AlaAlaProGlyglYTPGluGlyGlnAlaValaPheleuSerPheAspGlyglY 120
Db 314 GCTAGGCGAGAGGTGCTGGAGAGGCGAGAACTTTAACTGACTTTCAACAGCGGA 373
QY 121 GlyAlaIleGluPheAlaGlnleuMetVallylaAlaSerAlaAlaAarglyIle 140
Db 374 GGAGCCATCGAGTTGACAGCTGATGTCAAGCTCCCTAGTGCTTCACTGAGGATT 433
QY 141 ProleuGlySerValleuTyrtPheAspThrSerGlyleuTyrtlleIleThrValPro 160

Db 434 CTTCTC----- 439
QY 161 GlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrProIleValIleTyrgly 180
Db 439 ----- 439
QY 181 ProProProGlyIlyThrValGlnProGlyGluTyrglyThrProProGluGlyTyr 200
Db 440 ---CAAGCCCTGAGCTAT-----GATATC 460
QY 201 GlyAlaGlnProGlyIlyTyrglyAlaProPromeGlyTyrglyAlaProProValGly 220
Db 461 ACACTGATACCTGAGAGGAT---GCACCT-----GTCCACCTGCT--- 499
QY 221 TyrglyValaProProGlyIlyTyrglyValaProProGlyIlyTyrglyValaProProGly 240
Db 500 -----CCAGAGGGGTATGACCTGCTCCAGAGGGATATCACCTGCTCGGGG 547
QY 241 GlyTyrglyAlaProProGlyIlyTyrglyValaProProGlyIlyTyrglyAlaProPro 260
Db 548 GAGCTATGCTCTCTCTCA-----CCACCTCAAAAGGT-----CCT 583
QY 261 GlyIlyTyrglyAlaProProAla---GlyTyrglyAlaProPro 274
Db 584 TATCTTATGACACCACTCGAATGATGCTACGAGCTGCTCC 628

RESULT 12
BQ735296 862 bp mRNA linear EST 16-JUL-2002
LOCUS BQ735296
DEFINITION AGENCOUNT 8096274 NICHD_XGC_Bmd4 Xenopus laevis cDNA clone
IMAG:554398 5', mRNA sequence.
ACCESSION BQ735296
VERSION BQ735296.1 GI:21874193
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
1 (bases 1 to 862)
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
TITLE Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Igor David
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMLT)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LMLT at:
www.bio.lml.gov/bdrr/image/image.html
Plate: LLM12243 row: c column: 21
High quality sequence stop: 710.

FEATURES
source
1. 862
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/dev_stage="embryo, stage 31-32"
/lab_host="DH10B (phage-resistant)"
/clone="NICHD_XGC_Bmd4"
/note="Organ: whole embryo; Vector: pCMV-SPORT6; Site: 1;
Nori; Site 2: Salt cloned unidirectionally. Primer: Oligo
dT. Average insert size 2.1 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection (XGC
Library)." 1 library.

BASE COUNT 233 a 247 c 186 g 195 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 3.61e-28 Length: 862
 Score: 536.00 Matches: 121
 Percent Similarity: 54.32% Conservative: 30
 Best Local Similarity: 43.53% Mismatches: 71
 Query Match: 31.36% Indels: 56
 DB: 13 Gaps: 10

US-09-864-291-5 (1-313) x BQ735296 (1-862)

Qy 1 MetAlaValaAngInSerHlethrcIuSerArGtAglYAlaLeuIleProSerGlyGlu 20
 Db 21 ATGTCACTTAATAAGAACCACTCACAGGG---GCCGGGATCATGTGTACAGCAAGCGGAG 77
 Qy 21 SerValLeuLyGInCyGluAspValAspLeuCySphLeuGlnLysProValGluSer 40
 Db 78 AGTATTCTCAGACAGGTAAAGATGTGAACTCTCTCAGGATATATGTCTCACAAATCA 137
 Qy 41 TyrLeuPheAenGlyThrlYleYleGlyThrlPheLeuThrSerTyrArgValAlaPhe 60
 Db 138 GAAGCATTCAAGGCAAAAGAAAGATCGCTATATCTCACCGCTTACAGGGGTATCTTT 197
 Qy 61 ValThrSerHleuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 Db 198 CTAAGCAAGG-----AAAGACCCCATGATGTCTCTCATATGCCATTCTACCTGATG 251
 Qy 81 SerAspCySthrIleGluGlnProIlePheAlaProAsnTyrIleYleGlyThrlGln 100
 Db 252 AAAGATGCTCATATGAACAGCCAGCTCTCTGCACTACATCAAGGAAACCATCACT 311
 Qy 101 AlaAlaProGlyGlyGlyTrrGlnGlyGlnAlaValAlaPheLysLeuSerPheArgGly 120
 Db 312 GCTGAACCTCGAGATGCTGGAGGGGTAAACCTCTTTAACTTAACCTTAACAGTGA 371
 Qy 121 GlyAlaIleGluPheAlaGlnLeuMetValLysAlaAlaSerAlaAlaArgGlyTyr 140
 Db 372 GGAAGTATTGATTTGACAAATCATGTTCAAAATGCAACTTCTGCTTCAGACCCCT 431
 Qy 141 ProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleThrValPro 160
 Db 432 CCT-----GTTCCT 440
 Qy 161 GlyAlaAlaValCysSerSerGlnThrProCySproAlaTyrProIleValIleTyrGly 180
 Db 441 CATGCTGCATAT----- 452
 Qy 181 ProProProProGlyTyrThrValGlnProGlyGlyTyrGlyTyrProProGlyGlyTyr 200
 Db 453 -----GGATACACTCTCTGCTGCAAGAGATATGA-----CCAGGCGGGTAC 494
 Qy 201 GlyAlaGlnProGlyGlyTyrGlyValaProProMetGlyTyrGlyValaProProValGly 220
 Db 495 CCTCAGAGCTCAGGAATATATACACCCCAACCATCAAGG-----CCTATATCA 548
 Qy 221 TyrGlyValProPro---GlyGlyTyrGlyValaProProGlyGlyTyrGlyValaProPro 239
 Db 549 TATGAGACCCCTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 605
 Qy 240 GlyGlyTyrGlyValaProProGlyGlyTyrGlyValaProProGlyGlyTyrGlyValaPro 259
 Db 606 -----TATGCTCCGCAACAGGA-----CTGGAGATGATATCACCCCT 644
 Qy 260 ProGlyGly-----TyrGlyValaProProAlaGlyTyrGlyValaProPro 274
 Db 645 CCTGAATGAATCCCATTTATACATGCTCTCTCACACACCTTATCTGAGACCCCT 698

RESULT 13

CD301781

LOCUS 892 bp mRNA linear EST 27-MAY-2003

DEFINITION AGNCORRT 14238680 NICHD XGC Byel Xenopus laevis cDNA clone

IMAGE:6957790 5', mRNA sequence.

ACCESSION CD301781

VERSION CD301781.1

KEYWORDS EST.

SOURCE

ORGANISM

Xenopus laevis (African clawed frog)
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopus; Xenopus.

REFERENCE

1. (bases 1 to 892)

NIH-MGC <http://mgc.ncl.nih.gov/>

TITLE

JOURNAL

COMMENT

Unpublished
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: gsgerha@mail.nih.gov
 Tissue Procurement:
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: L14M14587 row: a column: 21
 High quality sequence stop: 646.
 Location/Qualifiers

FEATURES

source

1..892
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:6957790"
 /lab_host="adult"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NICHD_XGC_Byel"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site: 1; NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2.3 kb. Constructed by Life
 Technologies. Note: This is a Xenopus Gene Collection (XGC
) library."

BASE COUNT 239 a 247 c 198 g 208 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.26e-27 Length: 892
 Score: 528.50 Matches: 125
 Percent Similarity: 52.04% Conservative: 28
 Best Local Similarity: 42.52% Mismatches: 82
 Query Match: 30.92% Indels: 59
 DB: 14 Gaps: 11

US-09-864-291-5 (1-313) x CD301781 (1-892)

Qy 1 MetAlaValaAngInSerHlethrcIuSerArGtAglYAlaLeuIleProSerGlyGlu 20
 Db 41 ATGTCACTTAATAAGAACCACTCACAGGG---GCCGGGATCATGTGTACAGCAAGCGGAG 97
 Qy 21 SerValLeuLyGInCyGluAspValAspLeuCySphLeuGlnLysProValGluSer 40
 Db 98 AGTATTCTCAGACAGGTAAAGATGTGAACTCTCTCAGGATATATGTCTCACAAATCA 157
 Qy 41 TyrLeuPheAenGlyThrlYleYleGlyThrlPheLeuThrSerTyrArgValAlaPhe 60
 Db 158 GAAGCATTCAAGGCAAAAGAAAGATCGCTATATCTCACCGCTTACAGGGTATCTTC 217
 Qy 61 ValThrSerHleuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 Db 218 CTAAGCAAGG-----AAAGATCCCATGATGTCTTAAATATGCAATCTTACATGATG 271
 Qy 81 SerAspCySthrIleGluGlnProIlePheAlaProAsnTyrIleYleGlyThrlGln 100
 Db 272 AAAGATGCTCATATGAACAGCCAGGTCTCTGCACTACATCAAGGAAACCATCACT 331
 Qy 101 AlaAlaProGlyGlyGlyTrrGlnGlyGlnAlaValAlaPheLysLeuSerPheArgGly 120
 Db 332 GCTGAACCTCGAGAGGGCTGGAGGGTCAAGCTCTTTAAACTTACATTTAATAGTGA 391

QY 121 GYLAALIEGUPHEALAGLNUETVALYVALASERIALAALAEAGLYILE 140
 DB 392 GAGCTATTGAATTTGACAAATCATGTTCAAAATGACACTGTCTTCCAGAACCTT 451
 QY 141 ProeuGlySerValaenTyTrpPheapThrsSerGlyLeuTyTleTherValPro 160
 DB 452 CCT-----GTTCT 460
 QY 161 GYLAALAEVALCYSSERSEGLINThrProCySProAlaTyProIleValIleTyGly 180
 DB 461 CATGCTGCATAT----- 472
 QY 181 ProProProProGlyTyTrpThrValGlnProGlyGlyTyTrpThrProProGlyGlyTyTr 200
 DB 473 -----GATACACACCTGCTCCAGAGGCTATGCA-----CATGGCGGTAC 514
 QY 201 GYLAAGLInProGlyGlyTyTrpValAlaProProMetGlyTyTrpValAlaProProValGly 220
 DB 515 CCTCAGCTCCAGAAATATATACACCCCAACCCCAACATCAGGC-----CTTATCCA 568
 QY 221 TyTrpValProPro---GlyGlyTyTrpValProProGlyGlyTyTrpValProPro 239
 DB 569 TATGACACCCCTGCTATGATGATATGACCACTCCAAATCCATATGATACCCG--- 625
 QY 240 GlyGlyTyTrpValAlaProProGlyGlyTyTrpValProProGlyGlyTyTrpValAlaPro 259
 DB 626 -----TATGCTCCACAGCAGCA-----CCTGGCATATATCCACCCCT 664
 QY 260 ProGlyGly-----TyTrpValAlaProProAlaGlyTyTrpValAlaProProAlaGly 276
 DB 665 CTTGAATGAATATCCATTATACATGCTCTCTCCACCACTTATCTCGAAGCCCTTACAT 724
 QY 277 AaenGluAlaLeu-----ProProAlaTyTrpGluAlaPro 287
 DB 725 GGGAGCCCCCTAATGATGAGACCCCAACCTTTCGCTCCCA 766
 RESULT 14
 CB593075 719 bp mRNA linear EST 03-APR-2003
 LOCUS AGENCOURT 12918013 NICHD XGC Tel Xenopus laevis cDNA clone
 DEFINITION IMAGE:6899092 5', mRNA sequence.
 CB593075
 ACCESSION CB593075.1 GI:29510931
 VERSION EST.
 KEYWORDS Xenopus laevis (African clawed frog)
 ORGANISM Xenopus laevis
 SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 719)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@db-remail.nih.gov
 Tissue Procurement: Dr. Igor David, NICHD
 CDNA Library Preparation: CLONTECH
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNU)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNU at:
 http://image.llnl.gov
 Plate: LCM3240 row: h column: 03
 High quality sequence stop: 605.
 Location/Qualifiers
 1..719
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:6899092"
 /label_type="Normal testis"

/lab host="DH10B Tona"
 /clone lib="NICHD XGC Tel"
 /note="Vector: pDNR-LIB; Site 1: SfiI (directional);
 Site 2: SfiI (directional); Non-normalized full-length
 enriched library 5' and 3' adaptors used in cloning
 as follows: 5' adaptor sequence: 5'-CAGCGCATTAAGCC-3'
 and 3' adaptor sequence:
 5'-ATTCTAAGGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A, C,
 or G and N = A, C, G, or T). Average insert size 1.2kb
 (cdna range 0.4-3.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Corp."

BASIS COUNT 200 a 195 c 159 g 165 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6,366-27 Length: 719
 Score: 517.00 Matches: 121
 Percent Similarity: 54.32% Conservative: 30
 Best Local Similarity: 43.53% Mismatches: 71
 Query Match: 30.25% Indels: 57
 DB: 14 Gaps: 10

US-09-864-291-5 (1-313) x CB593075 (1-719)
 QY 1 MetAlaValaenGlnSerHisThrGluSerArgGlyValaLeuIleProSerGlyGlu 20
 DB 39 ATGCACTTAAATGAACCACTCAGAGGG---GCCGATC-ATTGACACCAAGGGAG 94
 QY 21 SerValLeuGlnGlyGluAapValaPheLeuCySPheLeuGlnTySProValGluSer 40
 DB 95 AGTATCTCAGACAGTAAAGATGAGAACTCTCTCAGCAATATGCTCACAATCA 154
 QY 41 TyTrpPheAenGlyThrTyLeuGlyGlyThrPheLeuThrsTyTrpValAlaPhe 60
 DB 155 GAAGCATTCAGAGCAGAAAGAAAGAGTCCATATCTCACCCGCTACAGGGTATCTTT 214
 QY 61 ValThrsSerHisLeuValaAapProMetLeuSerPheMetCProPheGlyLeuMet 80
 DB 215 CTAGAGCAAGAGG-----AAGAACCCCATATGCTTCAATATGCAATTCATG 268
 QY 81 SerPheCySPheThrIleGlnGlnProIlePheAlaProAenTyTleTyGlyThrIleGln 100
 DB 269 AAGAGATGCTCATAGAACAGCCAGCTTCTGCAACATCAAGGAGCAATCATG 328
 QY 101 AlAlaAlaProGlyGlyTyTrpGluGlnAlaValaPheTyLeuSerPheAlaGlyGly 120
 DB 329 GCTGAACCTGGAATGCTGGAGGGTAAACCTCTTTTAACTTCAATTAACATGGA 388
 QY 121 GYLAALIEGUPHEALAGLNUETVALYVALASERIALAALAEAGLYILE 140
 DB 389 GAGGCTATTGAATTTGACAAATCATGTTCAAAATGCAACTTCTTCCAGAACCTT 448
 QY 141 ProeuGlySerValaenTyTrpPheapThrsSerGlyLeuTyTleTherValPro 160
 DB 449 CCT-----GTTCT 457
 QY 161 GYLAALAEVALCYSSERSEGLINThrProCySProAlaTyProIleValIleTyGly 180
 DB 458 CATGCTGCATAT----- 469
 QY 181 ProProProProGlyTyTrpThrValGlnProGlyGlyTyTrpThrProProGlyGlyTyTr 200
 DB 470 -----GATACACTCTGCTCCAGAGGCTATGCA-----CATGGCGGTAC 511
 QY 201 GYLAAGLInProGlyGlyTyTrpValAlaProProMetGlyTyTrpValAlaProProValGly 220
 DB 512 CCTCAGCTCCAGAAATATATACACCCCAACCCCAACATCAGGC-----CTTATCCA 565
 QY 221 TyTrpValProPro---GlyGlyTyTrpValProProGlyGlyTyTrpValProPro 239
 DB 566 TATGACACCCCTGCTATGATGATATGACCACTCCAAATGCTATGATACCCG--- 622

QY 240 GlyGlyTyrGlyAlaProProGlyGlyTyrGlyValProProGlyGlyTyrGlyAlaPro 259
 DB 623 -----TATGCTCCGCCACACAGG-----CTGGCATGATTCACACCCCT 661
 QY 260 ProGlyGly-----TyrGlyAlaProProAlaGlyTyrGlyAlaProPro 274
 DB 662 CCTGAATGAATCCATTTACATGCTCTCTCCACACCTTATCTGAGACCCCG 715
 RESULT 15
 AL898723
 LOCUS AL898723 XGC-egg Silurana tropicalis cDNA clone T89036a07 5', mRNA
 DEFINITION AL898723 XGC-egg Silurana tropicalis EST project 2002
 ACCESSION AL898723
 VERSION AL898723.1 GI:22951115
 KEYWORDS EST
 SOURCE Silurana tropicalis (western clawed frog)
 ORGANISM Silurana tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Silurana.
 1 (bases 1 to 655)
 Taylor, R., Ashurst, J. L., Croning, M. D. R., Zorn, A. M. and Rogers, J.
 Sanger Xenopus tropicalis EST project 2002
 JOURNAL Unpublished
 COMMENT Contact: Taylor R
 Sanger Centre
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS SEQUENCE ID: T89036a07.plkSP6
 Sequencing primer: SP6
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Aaron M. Zorn.
 FEATURES
 source
 1..655
 Location/Qualifiers
 /organism="Silurana tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="T89036a07"
 /dev_stage="egg"
 /lab_host="Escherichia coli XL1-blue"
 /clone_1lb="XGC-egg"
 /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
 was oligo dt primed from 5ug of poly A+ RNA from egg.
 EcoRI-NotI cut cDNA was then ligated into pCS107 with
 EcoRI at the 5' end and NotI at the 3' end"
 BASE COUNT 172 a 172 c 155 g 156 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.1e-26 Length: 655
 Score: 513.00 Matches: 112
 Percent Similarity: 55.82% Conservative: 27
 Best Local Similarity: 44.98% Mismatches: 62
 Query Match: 30.02% Indels: 48
 DB: 9 Gaps: 8
 US-09-864-291-5 (1-313) x AL898723 (1-655)
 QY 1 MetAlaValAlaSerGlnSerIleThrGluSerArgGlyAlaIleuIleProSerGlyGlu 20
 DB 39 ATGTCCCTAAATTAAGACCACTCTGAGGG---GCTGGATCATCTCTCAGCAACGGGAG 95
 QY 21 SerValLeuLysGlnCysGluAlaPValAlaPLeuCyPheLeuGlnLysProValGluSer 40
 DB 96 AGTGTCTCAGGACAGTGAAGATGTGGAATCTCTCTCAGCATATGTCTCAAAAATCA 155
 QY 41 TyrIleuPheAsnGlyThrIleuLysGlyTyrIleuPheLeuThrSerTyrArgValAlaPhe 60
 DB 156 GAGGATTCAAAGGACCAAAAAGATCATCTTTCTACACCCCTACAGGGATGATCTTC 215

QY 61 ValThrSerHisLeuValaAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 DB 216 CTAAGCAAGGGG-----AAGACCCATGATATGCTTTATATGACATTTATCTGATG 269
 QY 81 SerAspCysThrIleGluGlnProIlePheAlaProAsnTyrIleGlyTyrIleGln 100
 DB 270 AAAGATGCTTCATAGAACCCAGTGTCTCTCCACATCAATCAAGGATCAATGAGT 329
 QY 101 AlaAlaProGlyGlyGlyTyrGluGlyGlyAlaValaPheLysLeuSerPheArgLysGly 120
 DB 330 GCTAGGCTGAAGGTGGGTGGGTGGGTCAAGCCCTTTCAAACTTACTTAAATAGTGA 389
 QY 121 GlyAlaIleGluPheAlaGlnLeuMetValLysAlaAlaSerAlaAlaAlaArgLys 140
 DB 390 GGTCTATTGAATTGACAAAGTCATGTTCAAAATGCAACTGTGTCTTCCAGA----- 443
 QY 141 ProLeuGlySerValaAsnTyrTrpPheAspPheSerGlyLeuTyrIleIleThrValPro 160
 DB 443 ----- 443
 QY 161 GlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrProIleValIleTyrGly 180
 DB 444 -----GCCCTCTCTGCTCCTCATGCCGTATATGGA 473
 QY 181 -----ProProProProGlyTyrThrValGlnProGlyGlyTyrThrProProGlu 198
 DB 474 TACATTCCTGCACAGCGGGGTACCTCCAGCTCCAGGAATATATTCAGCCCAACCCCA 533
 QY 199 GlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProMet---GlyTyrGlyAlaPro 217
 DB 534 CCAATCAGGACCTTATCCA-----TACGAGCCCCCTGCTATGAATGATATGAGCACT 587
 QY 218 ProValGlyTyrGlyValProProGlyGlyTyrGlyAlaProProGlyGlyTyrGlyVal 237
 DB 588 CCAATCAGGATGATACCC-----TATGCTCAGCACACAGGA----- 626
 QY 238 ProProGlyGlyTyrGlyAlaProPro 246
 DB 627 ---CCTGGTATGATCCACCCCTCTCT 650

Search completed: December 16, 2003, 20:45:08
 Job time : 2048.43 secs